

RAPID MIXING FOR SAMPLING COLORINGS OF RANDOM GRAPHS UP TO TREE UNIQUENESS THRESHOLD

ZHIDAN LI, KUAN YANG

ABSTRACT. We study the mixing time of the single-site Glauber dynamics for sampling proper q -colorings on sparse Erdős-Rényi random graphs $\mathcal{G}(n, d/n)$. We prove that, whenever $q \geq (1 + o_d(1))d$, Glauber dynamics mixes in polynomial time with high probability over $\mathcal{G}(n, d/n)$ for all sufficiently large integer n . This improves the previous bound $q = (\alpha^* + o_d(1))d$ where $\alpha^* = 1.7632\dots$ is the unique solution of $x = \exp(1/x)$ on $[1, \infty)$ by Efthymiou, Hayes, Štefankovič, and Vigoda (SODA, 2018) and asymptotically matches the uniqueness threshold on infinite regular trees.

Our proof has three main ingredients. First we establish logarithmic bounds on the spectral and coupling independence constants of the Gibbs distribution by extending the tree analysis of Chen, Liu, Mani, and Moitra (J. ACM, 2026). Second, we extend the entropy-factorization framework for two-spin systems on random graphs, developed by Bezáková, Galanis, Goldberg, and Štefankovič (TALG, 2024), to proper colorings by introducing a buffer around high-degree vertices. Finally, we compare the resulting heat-bath block dynamics with the single-site Glauber dynamics.

(Zhidan Li, Kuan Yang) SCHOOL OF COMPUTER SCIENCE, SHANGHAI JIAO TONG UNIVERSITY, SHANGHAI, CHINA.
Email: yueEnTeRnAL@sjtu.edu.cn, kuan.yang@sjtu.edu.cn.

1. INTRODUCTION

The model of proper colorings, which is often viewed as the anti-ferromagnetic Potts models at zero temperature, plays a fundamental role in combinatorics, probability theory, statistical physics and theoretical computer science. Given a graph $G = (V, E)$ and a color number $q \geq 2$, a *proper q -coloring* is an assignment of colors from $[q] := \{1, \dots, q\}$ to the vertices of G such that adjacent vertices receive different colors. Let $\Omega_{G,q}$ denote the set of proper q -colorings of G , and let $\mu_{G,q}$ be the uniform distribution on $\Omega_{G,q}$. Efficient sampling from $\mu_{G,q}$ is a basic and important algorithmic problem, closely related to approximate counting and the study of Gibbs measures. This problem is hard even on triangle-free graphs $\Delta > q$ for even q , unless $\mathbf{NP} = \mathbf{RP}$ ¹ [GŠV15].

In this paper, we study sampling proper colorings of *sparse Erdős–Rényi random graphs*. For a fixed real number $d > 1$, the random graph $\mathcal{G}(n, d/n)$ has vertex set $[n]$ and includes each pair in $\binom{[n]}{2}$ independently with probability d/n . Its average degree converges to d with high probability, whereas its maximum degree is $\Theta(\ln n / \ln \ln n)$ with high probability. Thus, the maximum degree and the typical local branching rate are on very different scales. This distinction makes it natural to seek sampling algorithms whose color requirement depends on the average degree d rather than on the maximum degree.

The statistical physics of random graph colorings provides several benchmarks for this problem. For large q , the cavity method predicts a *clustering* transition at average degree of order $q \ln q$, followed by condensation and, eventually, a loss of colorability near $2q \ln q$ [ZK07, BS20]. These transitions describe the global geometry of the space of proper colorings. By contrast, *Gibbs uniqueness* concerns the sensitivity of local marginals to distant boundary conditions and provides a natural benchmark for the behavior of local Markov chains. For the infinite Δ -regular tree, the Gibbs uniqueness threshold for proper colorings is $q \geq \Delta + 1$ (for sufficiently large integer Δ) [Jon02]. Related local-stability calculations for sparse Erdős–Rényi random graphs also suggest the bound $q = d + 1$ as a natural uniqueness threshold [ZK07]. These observations motivate the question of whether a local-update Markov chain, and in particular the natural *single-site Glauber dynamics*, can mix rapidly when the number of colors and the average degree are in the tree-uniqueness regime.

Over the past several decades, substantial progress has been made on almost uniform sampling proper colorings of sparse random graphs with a number of colors depending on the average degree rather than the maximum degree. Mossel and Sly [MS10] established rapid mixing of a Gibbs sampler when $q = q(d)$ is sufficiently large. Efthymiou [Eft14] subsequently introduced a weighted block decomposition that reduced the color requirement to a quantity linear in d . Combining block decomposition with local uniformity properties of the Gibbs distribution, Efthymiou, Hayes, Štefankovič, and Vigoda [EHŠV18] proved that the single-site Glauber dynamics mixes in polynomial time with high probability whenever $q \geq (\alpha^* + o_d(1))d$, where $\alpha^* = 1.7632\dots$ is the unique solution of $x = e^{1/x}$ on $[1, \infty)$.

However, a different line of work has reached a smaller color requirement if we relax the requirements of accuracy. Efthymiou [Eft16] developed a polynomial-time non-MCMC algorithm which, for $q \geq (1 + o_d(1))d$, produces a proper coloring whose distribution is within $n^{-\Omega(1)}$ total variation distance of $\mu_{G,q}$, with high probability over the random graph. Related deterministic work [Eft20] provides a $(1 \pm n^{-\Omega(1)})$ -approximation to the free energy of the coloring model in the same regime. However, the sampling accuracy in [Eft16] (and the counting accuracy in [Eft20]) is tied to the instance size, and these results do not establish polynomial-time mixing of the single-site Glauber dynamics or yield an **FPAUS** with an arbitrarily prescribed accuracy parameter. Consequently, a fundamental question remained open:

*Does there exist an **FPAUS** for proper q -colorings on $\mathcal{G}(n, d/n)$ when $q \geq (1 + o(1))d$?*

¹The hardness in [GŠV15] is demonstrated for the counting task, and thus for the sampling task by self-reducibility of list-coloring instances.

In particular, does the single-site Glauber dynamics mix in polynomial time?

We answer this question affirmatively, improving the previous rapid-mixing condition $q \geq (\alpha^* + o_d(1))d$ to $q \geq (1 + o_d(1))d$ and thereby asymptotically reaching the uniqueness threshold of the corresponding regular tree.

Theorem 1.1. *Fix $\eta \in (0, 1)$. For sufficiently large constants $d > 1$ and every integer $q \geq (1 + \eta)d$, with probability $1 - o_n(1)$ for $G \sim \mathcal{G}(n, d/n)$, the single-site Glauber dynamics for proper q -colorings of G has mixing time*

$$\tau_{\text{mix}}(P^{\text{GD}}, \varepsilon) \leq n^C (1 + \ln(1/\varepsilon)) \quad \text{for every } \varepsilon \in (0, 1),$$

where $C = C(\eta, d, q) > 0$ is a constant independent of n and ε .

Theorem 1.1 closes the asymptotic gap between the previously known rapid-mixing bound for Glauber dynamics on sparse random graphs and the tree uniqueness threshold. In particular, it shows that the presence of vertices whose degrees grow with n does not preclude polynomial-time mixing when the number of colors exceeds the average degree by an arbitrarily small fixed fraction and d is sufficiently large.

1.1. Proof overview. The proof of Theorem 1.1 has three main ingredients: a logarithmic bound on coupling and spectral independence, an entropy-factorization argument using a buffer around high-degree vertices, and a comparison between heat-bath block dynamics and the original single-site Glauber dynamics.

First step: spectral and coupling independence. Our starting point is *spectral independence*, a framework for translating bounds on conditional influences into rapid mixing.

Definition 1.2 (Influence matrix and spectral independence [CGŠV21]). Given a distribution μ over $\Omega \subseteq [q]^V$, for every boundary $\Lambda \subseteq V$ and feasible boundary condition $\tau \in [q]^\Lambda$ (i.e., $\mu(\cdot \mid \tau)$ is well-defined), its influence matrix is defined as for all feasible $u \leftarrow i$ and $v \leftarrow j$,

$$\Psi_\mu^\tau(u \leftarrow i, v \leftarrow j) := \mu(v \leftarrow j \mid u \leftarrow i, \tau) - \mu(v \leftarrow j \mid \tau)$$

with zero on all entries with $u = v$ ². We say μ is γ -spectrally independent if for every feasible $\Lambda \subseteq V$ and $\tau \in [q]^\Lambda$, $\lambda_{\max}(\Psi_\mu^\tau) \leq \gamma$.

Spectral independence was first introduced in [ALOG24], building on the framework of high-dimensional expanders [AL20], to establish rapid mixing for the hard-core model up to the uniqueness threshold. It was subsequently developed for proper colorings on bounded-degree triangle-free graphs [FGYZ22, CGŠV21] and applied to a variety of spin systems, yielding rapid or optimal mixing bounds and further local-to-global principles [CLV23, CLV26, JPV22, BCC⁺22, CFYZ22, CE22]. For sparse random graphs, [BGGŠ24] established spectral independence for certain two-spin systems, including the hard-core model, using a branching-value analysis, while [EF23] sharpened the bounds and obtained rapid mixing of Glauber dynamics. It still remains open how to establish the spectral independence for multi-spin systems on random graphs.

We address these challenges through *coupling independence*, introduced in [CZ23].

Definition 1.3 (Coupling independence [CZ23]). A distribution μ over $\Omega \subseteq [q]^V$ is said to be C -coupling independent if for every boundary $\Lambda \subseteq V$ and every feasible partial pinnings σ, τ on Λ which differ only at exactly one coordinate, it holds that

$$\mathcal{W}_1(\mu(\cdot \mid \sigma), \mu(\cdot \mid \tau)) \leq C$$

²An alternative definition for Ψ_μ^τ is introduced in [FGYZ22].

where

$$\mathcal{W}_1(\mu, \nu) := \inf_{\gamma \text{ is a coupling of } (\mu, \nu)} \mathbb{E}_{(X, Y) \sim \gamma} \left[\sum_{u \in V} \mathbb{1}[X(u) \neq Y(u)] \right]$$

is the *Wasserstein-1 distance* of μ and ν . Denote by $\text{CI}(\mu)$ the coupling independence of μ , i.e., the minimal C such that μ is C -coupling independent.

Coupling independence is closely related to spectral independence and can be more convenient to establish by recursive coupling arguments. Recent work has used this approach to study the Glauber dynamics for proper colorings on bounded-degree graphs and graphs with large girth [CLMM26, CF25].

To show Theorem 1.1, we establish the following technical theorem, which is the first result extending coupling independence of proper coloring models to sparse Erdős-Rényi random graphs.

Theorem 1.4 (Coupling independence). *Fix $\eta \in (0, 1)$. For sufficiently large positive real $d > 1$ and every integer $q \geq (1 + \eta)d$, with probability at least $1 - o_n(1)$ over the Erdős-Rényi random graph $G \sim \mathcal{G}(n, d/n)$ for sufficiently large integer n , it holds that*

$$\text{CI}(\mu_{G,q}) \leq C \ln n$$

where $C = C(\eta, d, q) > 0$ is a constant depending only on η, d and q .

As a consequence, we obtain a spectral-independence bound of the same order.

Corollary 1.5 (Spectral independence). *Fix $\eta \in (0, 1)$. For sufficiently large positive real $d > 1$ and every integer $q \geq (1 + \eta)d$, with probability at least $1 - o_n(1)$ over the Erdős-Rényi random graph $G \sim \mathcal{G}(n, d/n)$ for sufficiently large integer n , $\mu_{G,q}$ is $O_{\eta,d,q}(\ln n)$ -spectrally independent.*

The proof of Theorem 1.4 combines the tree-based coupling method of [CLMM26] with the local structure of sparse random graphs. Although neighborhoods of logarithmic radius are trees or unicyclic with high probability, high-degree vertices prevent a direct application of bounded-degree correlation-decay estimates. We overcome this difficulty by introducing a message-passing process on locally tree-like graphs whose nodes are marked as good or bad. Good nodes lie sufficiently far from high-degree vertices and enjoy strong spatial mixing and total influence decay inherited from bounded-degree trees; bad nodes are handled by a more conservative message recursion. A comparison with a suitably marked Galton–Watson tree then bounds these messages and yields $O(\ln n)$ coupling independence uniformly over all feasible boundary conditions.

Second step: high-degree buffer and entropic factorization. We next adapt the entropy-factorization framework of [BGGŠ24] from two-spin systems to proper colorings.

Turning logarithmic spectral independence into a polynomial mixing bound requires special care because the coloring model has hard constraints and the random graph contains vertices of high degree. We choose a fixed degree cutoff and introduce a one-step buffer consisting of all high-degree vertices and their neighbors. Letting U denote the remaining low-degree vertices, we study the marginal coloring distribution on U . The buffer ensures that every proper coloring of the induced subgraph on U extends to a proper coloring of the original graph; together with connected conditional supports, marginal boundedness, and spectral independence, this makes it possible to apply a uniform block entropy-factorization theorem in [CLV26, BGGŠ24].

To pass from block entropy factorization to a single-site inequality, we analyze the geometry of a randomly chosen block together with the buffered high-degree region. A percolation argument adapted to the buffer shows that, except for a sufficiently small probability over the random block, every connected component of the induced subgraph has size $O(\ln n)$. On such components, an elementary entropy tensorization bound incurs only a polynomial loss. Combining this local bound

with uniform block entropy factorization yields an approximate single-site entropy tensorization inequality with a polynomial factor, and hence polynomial relaxation time for the marginal Glauber dynamics on U .

Final step: Markov chain comparison. The mixing bound for the marginal distribution does not immediately imply rapid mixing of the original Glauber dynamics, since the high-degree buffer has been marginalized out. To recover the full coloring distribution, we introduce a heat-bath block dynamics in which each update resamples one vertex of U together with the buffer. An orthogonal decomposition of $L^2(\mu_{G,q})$ relates the spectral gap of this block dynamics to that of the marginal Glauber dynamics. We then use the logarithmic size of the relevant connected components to bound the relaxation times of the conditional single-site chains within each block. Finally, a comparison of Dirichlet forms transfers the polynomial relaxation-time bound to the original single-site Glauber dynamics and completes the proof of Theorem 1.1.

1.2. Discussion. We conclude the introduction with two directions suggested by our results.

Coloring random graphs with fewer colors. For large q , the clustering and colorability thresholds of sparse random graph colorings occur at average degrees of order $q \ln q$ [ZK07, AN05]. These phenomena suggest that the geometry of the coloring space may remain interesting well below the tree uniqueness threshold. A natural direction is to determine whether efficient sampling is possible with fewer colors, for example when $q = d - O(1)$ or even $q = \Theta(d/\ln d)$.

Random sampling vs. deterministic approximate counting. Theorem 1.1 gives an **FPAUS** for $\mu_{G,q}$ with high probability whenever $q \geq (1 + \eta)d$ for any fixed $\eta > 0$ and sufficiently large d . The coupling-independence analysis also applies to the feasible list-coloring instances satisfying the list-size slack condition introduced in Section 2. On the deterministic side, recent work [LY26] combines strong spatial mixing with a recursive linear-fractional-programming framework to obtain an **FPTAS** for counting proper colorings on sparse random graphs when $q \geq (2 + o_d(1))d$. It remains open whether sufficiently strong spatial mixing can be established in the regime $q \geq (1 + o_d(1))d$ and used to obtain deterministic approximation algorithms for $|\Omega_{G,q}|$ at the same threshold.

1.3. Organization of this paper. Section 2 introduces the notation and the family of feasible list-coloring instances used throughout the paper. In Section 3, we establish the structural properties of sparse Erdős–Rényi random graphs needed in the analysis. Section 4 proves Theorem 1.4 and Corollary 1.5 by combining the coupling method of [CLMM26] with recursive messages on marked Galton–Watson trees. Finally, we conclude Section 5 by entropy factorization and Markov chain comparison in Theorem 1.1.

AI disclosure. ChatGPT Pro 5.6 Sol and 6 Astra help to generate the proof of coupling independence after the authors provide the rough idea and related references. The authors manage to accomplish the proof idea of the rapid mixing from spectral independence with GPT’s help to handle the high-degree buffers. Heavy calculations are provided by agents and verified by authors. The authors rewrite the proof and take full responsibility for the paper.

2. NOTATIONS

Before all, we fix some notations here to help understand our paper. Let $\mathbb{R}$ be all real numbers, $\mathbb{R}_{>0}$ be all positive real numbers and $\mathbb{N}$ be all natural numbers. For a natural number $n \in \mathbb{N}$, denote by $[n]$ the set $\{1, \dots, n\}$. We use $\log$ to denote the logarithm function, e to denote the natural logarithm base and $\ln$ to denote the natural logarithm function. Let H_N be the N -th harmonic number, *i.e.*, $H_N = \sum_{j=1}^N \frac{1}{j}$. For an event $\mathcal{E}$, we also view it as a subset of the total universe and hence use $\bar{\mathcal{E}}$ to denote its complement. The indicator function of $\mathcal{E}$ is denoted by $\mathbb{1}[\mathcal{E}]$.

A graph G consists of a vertex set $V(G)$ and an edge set $E(G) \subseteq \binom{V(G)}{2}$. When the context is clear, we simply use V and E to denote $V(G)$ and $E(G)$ respectively. For a subset $S \subseteq V$, we use $N_G(S)$ to denote the neighborhood of S in G and set $N_G[S] := S \cup N_G(S)$. We use $\deg_G(\cdot)$ to denote the degree function of G . Let $\text{dist}_G(\cdot, \cdot)$ be the distance metric in G . Let $B_G(v, \ell)$ be the ℓ -radius ball centered at v in G and $S_G(v, \ell)$ be the ℓ -radius sphere at v in G , i.e.,

$$B_G(v, \ell) := \{u \in V(G) \mid \text{dist}_G(u, v) \leq \ell\}, \quad S_G(v, \ell) := \{u \in V(G) \mid \text{dist}_G(u, v) = \ell\}.$$

Furthermore, for a vertex subset $S \subseteq V$, let $B_G(S, \ell) = \cup_{v \in S} B_G(v, \ell)$ be the collection of vertices in distance ℓ from S . With abuse of notations, we sometimes use the vertex subset S to denote the induced subgraph $G[S]$ when G is clear in context.

Given a finite universe U and an assignment $\sigma : U \rightarrow [q]$, we use $\sigma(S)$ to denote the partial assignment of σ on S . For a subset $\Lambda \subseteq U$, the notation σ_Λ is abused to denote a partial assignment on Λ . When context is clear, we drop scripts and use the singleton itself to denote the set only containing it. Meanwhile, we sometimes use $\sigma(v)$ to denote the color of σ on v as well and use the term $v \leftarrow c$ to denote the event that the vertex v is assigned value c .

Proper colorings and list-coloring instances. For an original graph G and a color number $q \geq 3$ where G is q -colorable, a feasible list-coloring instance $\Phi = (G_\Phi, L_\Phi = \{L_\Phi(v)\}_{v \in V_\Phi})$ consists of an underlying graph G_Φ which is an induced subgraph of G from the vertex subset V_Φ and color lists $L_\Phi(v) \subseteq [q]$ for every $v \in V_\Phi$ such that

$$\forall v \in V_\Phi, \quad |L_\Phi(v)| - \deg_{G_\Phi}(v) \geq q - \deg_G(v).$$

Let $\mathcal{L}(G, q)$ collect all feasible list-coloring instances of G and q , and μ_Φ be the uniform distribution of all proper list-colorings in Φ .

Probability theory. We use standard notations for probability theory. For simplicity, when the context is clear, we abuse the usage of a random variable and its law. For an event $\mathcal{E}$, we use the term ' $\cdot \mid \mathcal{E}$ ' to denote random variables conditional on $\mathcal{E}$. Let μ be a probability distribution on $\Omega \subseteq [q]^U$ where q is the number of colors and U is a finite universe. A partial assignment τ on a boundary $\Lambda \subseteq U$ is feasible if it can be extended to an assignment on U with positive mass in μ . We use μ^τ to denote the probability distribution induced by μ conditional on pinning τ on Λ and μ_S^τ to denote the marginal of μ^τ projected onto S . For two random variables X, Y , we say X is dominated by Y (denoted by $X \preceq Y$) if there exists a coupling γ of (X, Y) such that

$$\mathbb{P}_{(X, Y) \sim \gamma}(X \leq Y) = 1.$$

For $p \geq 1$, define the ℓ_p -norm of a random variable X as

$$\|X\|_p := \mathbb{E}[|X|^p]^{1/p}.$$

3. RANDOM GRAPH PROPERTIES

Before proving our main results Theorems 1.1 and 1.4, we investigate some structural properties of random graphs. These properties provide some intuitions on sparse random graphs $\mathcal{G}(n, d/n)$, which play crucial roles in our analysis.

3.1. Locality of random graphs. In this part, we consider the locality of random graphs. Let $\mathcal{E}_\ell^{\text{local}}$ be the event that for every vertex $v \in V$, the ball $B_G(v, \ell)$ is tree or unicyclic. Set

$$L := \left\lfloor \frac{\ln n}{1000 \ln(2ed)} \right\rfloor.$$

The following lemma shows that the event $\mathcal{E}_\ell^{\text{local}}$ holds with high probability for some $\ell = O(L)$.

Lemma 3.1. *For sufficiently large positive integer n , it holds that*

$$\mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \mathcal{E}_{5L}^{\text{local}}) \geq 1 - o_n(1).$$

We firstly provide the following probability of the occurrence of a subgraph containing two small cycles.

Lemma 3.2. *Let $\mathcal{E}_m^{\text{two}}$ be the event that G contains a connected subgraph on at most m vertices with at least two surplus edges. Then for $d > 1$, it holds that*

$$\mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \mathcal{E}_m^{\text{two}}) \leq \frac{ed^2}{n(ed-1)} m^2 (ed)^m.$$

Proof. Assume an arbitrary order on vertices and an arbitrary order on all possible edges. Fix $k = 1, \dots, m$. For every connected subgraph on k vertices with at least $k+1$ edges, we construct a witness as: pick a minimal spanning tree firstly and then choose two minimal surplus edges. To count the number of such subgraphs, it suffices to count the number of all possible witnesses. By the Cayley's formula and the bound $\binom{n}{k} \leq (en/k)^k$, since there are at most k^4 possible choices of two surplus edges in a tree of size k , the number of possible witnesses on k vertices in G is at most

$$\binom{n}{k} k^{k-2} k^4 \leq \frac{e^k n^k}{k^k} \cdot k^{k+2} = e^k n^k k^2.$$

Therefore

$$\begin{aligned} \mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \mathcal{E}_m^{\text{two}}) &\leq \sum_{k=1}^m e^k n^k k^2 \cdot \left(\frac{d}{n}\right)^{k+1} \\ &= \sum_{k=1}^m \frac{d}{n} (ed)^k k^2 \\ &\leq \frac{ed^2}{n(ed-1)} m^2 (ed)^m \end{aligned}$$

which concludes the lemma. $\square$

Proof of Lemma 3.1. Suppose that there exists a vertex $v \in V$ such that the $5L$ -radius ball $B_G(v, 5L)$ contains two cycles. Take a breath-first spanning tree of the ball and two edges $e_1 = (u, w), e_2 = (x, y)$ (with possibly repeated occurrence of a vertex) outside this spanning tree. Consider the subgraph generated by joining tree paths from the endpoints u, w, x and y to v in addition to e_1 and e_2 , i.e.,

$$v \xrightarrow{\text{tree path}} u \xrightarrow{e_1} w \xrightarrow{\text{tree path}} v \xrightarrow{\text{tree path}} x \xrightarrow{e_2} y \xrightarrow{\text{tree path}} v.$$

This subgraph is on at most $20L + 1$ vertices and contains two cycles. Then it holds that

$$\mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \overline{\mathcal{E}_{5L}^{\text{local}}}) \leq \mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \mathcal{E}_{20L+1}^{\text{two}}).$$

Apply Lemma 3.2 with $m = 20L + 1$, and we obtain that

$$\mathbb{P}_{G \sim \mathcal{G}(n, d/n)} (G \in \mathcal{E}_{5L}^{\text{local}}) \geq 1 - 2ed^2 \left(1 + \frac{\ln n}{50 \ln(2ed)}\right)^2 n^{-49/50} \geq 1 - O_d(n^{-49/50} (\ln n)^2)$$

which is $1 - o_n(1)$ for every sufficiently large integer n . $\square$

3.2. Small occupancy of high-degree vertices in long paths. The next property is that, for every long paths in $\mathcal{G}(n, d/n)$, although there might be some vertices with high degree in the path, they account for a small proportion.

For $\eta \in (0, 1)$ and $d > 1$, pick the cut-off D as

$$D := \lfloor (1 + \eta/3)d \rfloor.$$

The high-degree vertex set is defined as

$$V^{\text{HD}} := \{v \in V \mid \deg_G(v) > D\}.$$

Let $\mathcal{E}_{\ell, r}^{\text{good}}$ be the event that for every simple path in $\mathcal{G}(n, d/n)$ of on ℓ vertices (of length $\ell - 1$, equivalently), there are at least $\ell/2$ vertices outside $B_G(V^{\text{HD}}, r)$.

Lemma 3.3. *On the event $\mathcal{E}_{5L}^{\text{local}}$, for sufficiently large $d > 1$ and $r = O_\eta((\ln d)^2)$, it holds that*

$$\mathbb{P}_{G \sim \mathcal{G}(n, d/n)} \left(G \in \mathcal{E}_{\lfloor L/4 \rfloor, r}^{\text{good}} \mid \mathcal{E}_{5L}^{\text{local}} \right) \geq 1 - o_n(1).$$

To prove Lemma 3.3, we first need the following bound.

Lemma 3.4. *Fix an arbitrary graph J on a subset of $V = [n]$. For every subset $S \subseteq V$ and $z > 0$, it holds that*

$$\mathbb{P}(S \subseteq V^{\text{HD}} \mid J \subseteq G) \leq \exp \left(-zD|S| + 2z|E(J)| + \frac{|S|d}{2}(e^{2z} - 1) \right).$$

Proof. We write

$$\sum_{v \in S} \deg_G(v) = N(S, J) + \sum_{e=(u,v) \notin E(J)} |\{u, v\} \cap S| \cdot \mathbb{1}[e \in E(G)]$$

where $N(S, J)$ is the total number of degrees contributing to $\sum_{v \in S} \deg_G(v)$ in J . It holds that $N(S, J) \leq 2|E(J)|$ and

$$\sum_{e=(u,v) \notin E(J)} |\{u, v\} \cap S| \leq n|S|.$$

By the inequality $e^{zx} - 1 \leq \frac{x}{2}(e^{2z} - 1)$ for every $x \in [0, 2]$ and $z > 0$,

$$\begin{aligned} \mathbb{E} \left[\exp \left(z \sum_{e=(u,v) \notin E(J)} |\{u, v\} \cap S| \cdot \mathbb{1}[e \in E(G)] \right) \right] &= \prod_{e=(u,v) \notin E(J)} \left(1 + \frac{d}{n} (e^{z|\{u,v\} \cap S|} - 1) \right) \\ &\leq \exp \left(\frac{d}{n} \sum_{e=(u,v) \notin E(J)} (e^{z|\{u,v\} \cap S|} - 1) \right) \\ &\leq \exp \left(\frac{|S|d}{2} (e^{2z} - 1) \right). \end{aligned}$$

When $S \subseteq V^{\text{HD}}$, it must hold that $\sum_{v \in S} \deg_G(v) \geq D|S|$. Therefore by the Markov inequality,

$$\begin{aligned} \mathbb{P}(S \subseteq V^{\text{HD}} \mid J \subseteq G) &\leq \mathbb{P} \left(\sum_{e=(u,v) \notin E(J)} |\{u, v\} \cap S| \mathbb{1}[e \in E(G)] \geq D|S| - 2|E(J)| \mid J \subseteq G \right) \\ &\leq \exp \left(-zD|S| + 2z|E(J)| + \frac{|S|d}{2} (e^{2z} - 1) \right). \end{aligned}$$

Then we conclude the lemma. $\square$

Proof of Lemma 3.3. Set $L_0 = \lfloor L/4 \rfloor$. Suppose that there is a path $P = (u = v_0, \dots, v_{L_0-1})$ containing fewer than $L_0/2$ vertices outside $B_G(V^{\text{HD}}, r)$. Note that $B_G(P, r)$ is contained in $B_G(u, L_0 + r - 1)$. Therefore $B_G(u, L_0 + r - 1)$ is tree or unicyclic conditional on the event $\mathcal{E}_{5L}^{\text{local}}$ for all sufficiently large n . For every $v \in P \cap B_G(V^{\text{HD}}, r)$, pick a vertex $w = w(v) \in V^{\text{HD}}$ minimizing the distance $\text{dist}_G(v, w)$. Let

$$S := \{w = w(v) : v \in P \cap B_G(V^{\text{HD}}, r)\}.$$

Using a fact that in a tree or unicyclic graph, there are at most $4r + 2$ vertices in a simple path at distance $\leq r$ from a fixed vertex, it holds that

$$\frac{L_0}{2(4r + 2)} \leq |S| \leq L_0.$$

To apply Lemma 3.4, the graph J is constructed as the union of P and all shortest paths from v to $w(v)$ for $v \in P \cap B_G(V^{\text{HD}}, r)$. It is clear that J is connected and part of $B_G(u, L_0 + r - 1)$, meaning that J is a tree or a unicyclic graph. Thus

$$|V(J)| \leq (r + 1)L_0, \quad |E(J)| \leq |V(J)|, \quad |E(J)| \leq (r + 1)(8r + 4)|S|.$$

Apply Lemma 3.4 on J , S and $z = \frac{1}{2} \ln \frac{D - 8(r+1)(2r+1)}{d}$. It holds that for sufficiently large $d > 1$, $D - 8(r + 1)(2r + 1) \geq (1 + \eta/4)d$ and thus

$$\begin{aligned} \mathbb{P}(S \subseteq V^{\text{HD}} \mid J \subseteq G) &\leq \exp\left(-\frac{|S|d}{2}((1 + \eta/4) \ln(1 + \eta/4) - (1 + \eta/4) + 1)\right) \\ &\leq \exp\left(-\frac{C_\eta d L_0}{8r + 4}\right) \end{aligned}$$

where $C_\eta = \frac{1}{2}((1 + \eta/4) \ln(1 + \eta/4) - \eta/4) > 0$.

To conclude the lemma, we count the number of all possible J . Assume that J is on m vertices. Since J is tree or unicyclic, the expected number of all possible J is at most

$$\binom{n}{m} m^{m-2} \frac{d^{m-1}}{n^{m-1}} + \binom{n}{m} m^{m-2} \frac{m^2}{2} \frac{d^m}{n^m} \leq \left(\frac{n}{dm^2} + \frac{1}{2}\right) (ed)^m.$$

For a fixed J on m vertices, there are at most 2^m possible choices of S . Therefore, the failure probability is upper bounded by

$$\begin{aligned} \mathbb{P}_{G \sim \mathcal{G}(n, d/n)}(G \in \overline{\mathcal{E}_{L_0, r}^{\text{good}}}) &\leq \sum_{m=1}^{(r+1)L_0} \left(\frac{n}{dm^2} + \frac{1}{2}\right) (ed)^m \cdot 2^m \cdot \exp\left(-\frac{C_\eta d L_0}{8r + 4}\right) \\ &\leq n \left(\frac{1}{d} + \frac{1}{2}\right) \frac{2ed}{2ed - 1} (2ed)^{(r+1)L_0} \exp\left(-\frac{C_\eta d L_0}{8r + 4}\right) \\ &\leq n \left(\frac{1}{d} + \frac{1}{2}\right) \frac{2ed}{2ed - 1} \exp\left(-\left(\frac{C_\eta d}{8r + 4} - (r + 1) \ln(2ed)\right) L_0\right). \end{aligned}$$

Increasing d such that $\frac{C_\eta d}{8r + 4} - (r + 1) \ln(2ed) \geq 20000 \ln(2ed)$, the failure probability is at most $n^{-2+o(1)}$, which concludes the lemma. $\square$

4. COUPLING INDEPENDENCE FOR RANDOM GRAPH COLORINGS

In this section, we present the proof of Theorem 1.4. First, we introduce the result of strong spatial mixing and total influence decay for colorings on trees in [CLMM26]. Then we show how to analyze the coupling discrepancy on random graphs using the introduced result. Roughly speaking, a message passing process on local trees with good/bad marks would gives an upper bound to the

coupling independence under any arbitrary pinning. Lastly, we estimate relevant messages, which finally concludes Theorem 1.4.

4.1. Strong spatial mixing and total influence decay on trees. Before we analyze the coupling independence on $G \sim \mathcal{G}(n, d/n)$, let us recall the strong spatial mixing and total influence decay for colorings on trees derived by [CLMM26]. We consider the following form of strong spatial mixing and total influence decay.

Definition 4.1 (Strong spatial mixing [CLMM26]). The Gibbs distribution μ of a feasible list-coloring instance $\Phi = (G = (V, E), L = \{L(v)\}_{v \in V})$ exhibits *strong spatial mixing* with exponential decay rate $1 - \delta$ and constant $C_{\text{SSM}} > 0$ if the following holds. For every integer $\ell \geq 1$, every feasible partial pinning π on $\Lambda \subseteq V$, every vertex $v \in V \setminus \Lambda$, every $c \in L(v)$ and every two feasible partial colorings σ, τ on $S_G(v, \ell) \setminus \Lambda$,

$$\left| \frac{\mu(v \leftarrow c \mid \pi, \sigma)}{\mu(v \leftarrow c \mid \pi, \tau)} - 1 \right| \leq C_{\text{SSM}}(1 - \delta)^\ell.$$

Definition 4.2 (Total influence decay [CLMM26]). The Gibbs distribution μ of a feasible list-coloring instance $\Phi = (G = (V, E), L = \{L(v)\}_{v \in V})$ exhibits *total influence decay* with exponential decay rate $1 - \delta$ and constant $C_{\text{TID}} > 0$ if the following holds. For every integer $\ell \geq 1$, every feasible partial pinning π on $\Lambda \subseteq V$, every vertex $v \in V \setminus \Lambda$, every two colors $a, b \in L(v)$,

$$\sum_{u \in S_G(v, \ell) \setminus \Lambda} \|\mu(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq C_{\text{TID}}(1 - \delta)^\ell.$$

The following proposition describe the strong spatial mixing and total influence decay for proper colorings on trees.

Proposition 4.3 ([CLMM26, Proposition 8.8]). *If $q = (1 + \eta)\Delta$ for $\eta \in (0, 1)$ and $\Delta \geq 7/\eta^2$, for all trees of maximum degree Δ , the uniform distribution of proper q -colorings exhibits:*

- *strong spatial mixing with exponential decay rate $e^{-\eta/4}$ and constant*

$$C_{\text{SSM}} = \frac{2q}{(\eta\Delta)^{1/2}} e^{1/\eta + \eta/4} = O_\eta(\text{poly}(\Delta)).$$

- *total influence decay with exponential decay rate $e^{-\eta/4}$ and constant*

$$C_{\text{TID}} = \frac{q^3}{(\eta\Delta)^{3/2}(q-1)} e^{3/\eta} = O_\eta(\text{poly}(\Delta)).$$

The same rates of strong spatial mixing and total influence decay hold for every feasible list-coloring instance on a tree of maximum degree at most Δ satisfying

$$|L(v)| - \deg_T(v) \geq q - \Delta \quad \text{for every free vertex } v \in V(T).$$

This follows from the pinned-leaf reduction recorded in Lemma A.1.

An application of Proposition 4.3 helps us analyze the coupling discrepancy on local parts of a global graph.

Corollary 4.4. *For every fixed parameter $\eta \in (0, 1)$, there exist $\kappa = \kappa(\eta) > 0, \rho \in (0, 1)$ and integers*

$$R = O_\eta(\ln d), \quad r = O_\eta((\ln d)^2) > R + 2$$

such that

$$(1) \quad 4R \cdot e^{-\kappa \cdot R} \cdot \rho^{-R} H_{D^R} \leq 1$$

and the following holds for all sufficiently large $d > 1$. Let $\Phi \in \mathcal{L}(G, q)$ and suppose that the r -radius ball of v is a tree and contains no vertex in V^{HD} . For every feasible partial pinning π on Λ in Φ and two feasible colors $a, b \in L_\Phi(v)$, it holds that

$$\sum_{\text{free vertex } u \in S_G(v, R)} \|\mu_\Phi(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu_\Phi(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq e^{-\kappa \cdot R}.$$

Proof. We first prove a weaker version restricted to trees by verifying the parameter choices in Proposition 4.3. Then we show how to remove the dependence on q in the weak results. Finally, we embed the refined SSM and TID results on small trees into global graph.

Define

$$\Delta := \max \{D, \lceil 2q/3 \rceil\}, \quad \delta = q/\Delta - 1.$$

For sufficiently large $d > 1$, it holds that

$$\frac{\eta}{4} \leq \min \left\{ \frac{2\eta}{3+\eta}, \frac{q-3}{2q+3} \right\} \leq \delta \leq \frac{1}{2}, \quad \Delta \geq 7/\delta^2.$$

We focus on the local tree at v of radius r :

$$T := G_\Phi[B_{G_\Phi}(v, r)].$$

After imposing the pinning π and restricting the instance to T , let T^π denote the forest induced by the free vertices and let $L^\pi(x)$ be the resulting effective list of each free vertex x . Observe that pinning and restriction to the local tree cannot decrease the gap between color lists and degree, *i.e.*, for every free vertex x ,

$$\begin{aligned} |L^\pi(x)| - \deg_{T^\pi}(x) &\geq |L_\Phi(x)| - \deg_{G_\Phi}(x) \\ &\geq q - \deg_G(x) \\ &\geq q - D \geq q - \Delta. \end{aligned}$$

Here the third inequality follows because the r -radius ball contains no vertex in V^{HD} , while the last inequality follows from $\Delta \geq D$.

Since T^π has maximum degree at most Δ and $q = (1 + \delta)\Delta$, every component of the induced list-coloring instance satisfies the hypotheses of Lemma A.1. Hence the SSM and TID bounds hold with exponential decay $e^{-\delta/4} \leq e^{-\eta/16}$ and constants

$$C_{\text{SSM}} = \frac{2q}{(\delta\Delta)^{1/2}} e^{1/\delta + \delta/4}, \quad C_{\text{TID}} = \frac{q^3 e^{3/\delta}}{(\delta\Delta)^{3/2}(q-1)}.$$

A further upper bound for C_{SSM} and C_{TID} is $B_\eta \sqrt{q}$ where

$$B_\eta := 4(6/\eta)^{3/2} e^{12/\eta}.$$

Therefore, it holds that for every two additional boundary pinning σ, τ differing at distance $k \geq 3$ of v ,

$$\left| \frac{\mu(v \leftarrow c \mid \pi, \sigma)}{\mu(v \leftarrow c \mid \pi, \tau)} - 1 \right| \leq B_\eta \sqrt{q} \cdot e^{-k \cdot \eta/16}$$

and for two feasible colors a, b at v and $k \geq 2$,

$$\sum_{\text{free vertex } u \in S_G(v, k)} \|\mu(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq B_\eta \sqrt{q} \cdot e^{-k \cdot \eta/16}.$$

where μ is the Gibbs distribution of proper list-colorings on T with all common pins and color lists.

To remove the dependence of q , we delete the vertex v and for every free child w , let T_w be the component of $T - v$ containing w . Let p_w be the marginal at w in T_w under the common pinning π . It holds that

$$\forall \text{ feasible color } c, \quad p_w(c) \leq \frac{1}{q - D + 1}$$

since w has at most $D - 1$ neighbors in T_w . Conditional on the assignment a or b at v , we only need to consider the maximal coupling of $p_w(\cdot \mid X_w \neq a)$ and $p_w(\cdot \mid X_w \neq b)$. The total variation distance is

$$\frac{\max \{p_w(a), p_w(b)\}}{1 - \min \{p_w(a), p_w(b)\}} \leq \frac{1/(q - D + 1)}{1 - 1/(q - D + 1)} = \frac{1}{q - D}.$$

Then we couple free vertices in $S_G(v, r)$ conditional on the assignment a or b at v as following. For every free child w of v , we maximally couple assignments at w as $p_w(\cdot \mid X_w \neq a)$ and $p_w(\cdot \mid X_w \neq b)$. When they agree, we use the identical coupling in the remaining vertices in T_w . Otherwise, we apply Proposition 4.3 at T_w . Therefore, the total variation distance at T_w is at most $\frac{1}{q-D} \cdot B_\eta \sqrt{q} \cdot e^{-(k-1)\eta/16}$. Summing over all children of v , the refined bound of total influence decay can be bounded by

$$\sum_{\text{free vertex } u \in S_G(v, k)} \|\mu(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq \frac{D}{q - D} B_\eta \sqrt{q} \cdot e^{-(k-1)\eta/16}.$$

Since $q \geq (1 + \eta)d$, $D \leq (1 + \eta/3)d$, and $x \mapsto \sqrt{x}/(x - D)$ is decreasing for $x > D$, we obtain that

$$\frac{D\sqrt{q}}{q - D} \leq \frac{3\sqrt{D}}{\eta}.$$

Therefore,

$$\sum_{\text{free } u \in S_G(v, k)} \|\mu(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq \frac{3B_\eta e^{\eta/16}}{\eta} \sqrt{D} e^{-k\eta/16}.$$

To refine the SSM, fix two additional pinnings σ, τ such that they differ only at distance at least $k \geq 3$ from v . For every free child w , let p_w^σ and p_w^τ be the marginals at w in T_w under the common pinning π and the additional pinnings σ, τ respectively. Since $q - D \geq 3$, it is clear to see p_w^σ and p_w^τ share a same support. Apply Proposition 4.3 at T_w and we obtain that for all feasible color c at w ,

$$|p_w^\sigma(c) - p_w^\tau(c)| \leq \frac{1}{q - D + 1} \cdot B_\eta \sqrt{q} \cdot e^{-(k-1)\eta/16}.$$

Set

$$W_c^\sigma = \prod_{w \text{ free child of } v} (1 - p_w^\sigma(c)), \quad W_c^\tau = \prod_{w \text{ free child of } v} (1 - p_w^\tau(c)).$$

By the mean value theorem,

$$\left| \ln \frac{W_c^\sigma}{W_c^\tau} \right| \leq \sum_{w \text{ free child of } v} \frac{|p_w^\sigma(c) - p_w^\tau(c)|}{1 - 1/(q - D + 1)} \leq \frac{D}{q - D} \cdot B_\eta \sqrt{q} \cdot e^{-(k-1)\eta/16}.$$

After normalization, we can obtain that

$$\left| \ln \frac{\mu(v \leftarrow c \mid \pi, \sigma)}{\mu(v \leftarrow c \mid \pi, \tau)} \right| \leq \frac{2D}{q - D} \cdot B_\eta \sqrt{q} \cdot e^{-(k-1)\eta/16}.$$

Note that $(1 - 1/(q - D + 1))^D \leq W_c^\sigma, W_c^\tau \leq 1$. Then it holds that

$$\left| \ln \frac{\mu(v \leftarrow c \mid \pi, \sigma)}{\mu(v \leftarrow c \mid \pi, \tau)} \right| \leq -2D \ln \left(1 - \frac{1}{q - D + 1} \right) \leq \frac{2D}{q - D} \leq \frac{6}{\eta}.$$

Therefore, by the inequality $|e^x - 1| \leq e^{|x|}|x|$, we have

$$\left| \frac{\mu(v \leftarrow c \mid \pi, \sigma)}{\mu(v \leftarrow c \mid \pi, \tau)} - 1 \right| \leq \frac{6\sqrt{D}e^{\eta/16+6/\eta}}{\eta} B_\eta \cdot e^{-k\cdot\eta/16}.$$

Hence, we have the refined SSM and TID exponential decay rate $e^{-\eta/16}$ and constant at most $C := 512\eta^{-3}e^{20/\eta} \cdot D^2$.

Now we show how to embed the SSM and TID results on trees into the global graph. Let $S_r := S_{G_\Phi}(v, r) \setminus \Lambda$ and $S_R := S_{G_\Phi}(v, R) \setminus \Lambda$ be the free spheres at radii r and R respectively. Let ν the law of X_{S_r} under the common pinning π . Then it holds that for every two feasible boundaries σ, τ on S_r ,

$$|\mu_\Phi(v \leftarrow c \mid \pi, \sigma) - \mu_\Phi(v \leftarrow c \mid \pi, \tau)| \leq \mu_\Phi(v \leftarrow c \mid \pi, \tau) \cdot Ce^{-r\cdot\eta/16}.$$

Averaging over τ with the law ν , it holds that

$$\left| \mu_\Phi(v \leftarrow c \mid \pi, \sigma) - \sum_{\tau} \nu(\tau) \mu_\Phi(v \leftarrow c \mid \pi, \tau) \right| \leq Ce^{-r\cdot\eta/16} \sum_{\tau} \nu(\tau) \mu_\Phi(v \leftarrow c \mid \pi, \tau).$$

Then by the Bayes' rule, it holds that

$$\|\nu(\cdot \mid v \leftarrow a) - \nu(\cdot \mid v \leftarrow b)\|_{\text{TV}} \leq \frac{C}{2} e^{-r\cdot\eta/16}$$

meaning that $\|\nu(\cdot \mid v \leftarrow a) - \nu(\cdot \mid v \leftarrow b)\|_{\text{TV}} \leq Ce^{-r\cdot\eta/16}$.

Therefore, we maximally couple $\sigma \sim \nu(\cdot \mid v \leftarrow a)$ and $\tau \sim \nu(\cdot \mid v \leftarrow b)$ and apply the refined TID results when $\sigma = \tau$. By the crude bound that the disagreements are at most $|S_R|$ when $\sigma \neq \tau$, it holds that

$$\sum_{\text{free vertex } u \in S_G(v, R)} \|\mu_\Phi(u \leftarrow \cdot \mid \pi, v \leftarrow a) - \mu_\Phi(u \leftarrow \cdot \mid \pi, v \leftarrow b)\|_{\text{TV}} \leq |S_R| Ce^{-r\cdot\eta/16} + Ce^{-R\cdot\eta/16}.$$

We pick $\kappa = \eta/32$, R as the smallest integer at least

$$\max \left\{ 4, \frac{32}{\eta} \ln(2C), \frac{8}{\kappa} \ln D \right\}$$

and r as the smallest integer at least

$$\max \left\{ R + 3, \frac{16R \ln D + 16\kappa R + 16 \ln(2C)}{\eta} \right\}.$$

It is clear to see $R = O_\eta(\ln d)$ and $r = O_\eta((\ln d)^2)$ and the desired total influence decay holds. Moreover, pick $\rho = e^{-\kappa/2}$. For sufficiently large $d > 1$, it can be easily verified that $4R \cdot e^{-\kappa \cdot R} \cdot \rho^{-R} H_{DR} \leq 1$. Thus we conclude the corollary. $\square$

We pick parameters κ, ρ, R and r with respect to η and d in the remaining part of Section 4 as Corollary 4.4.

The second tool we introduce from [CLMM26] is its *one-disagreement coupling*. We recall it here: to couple the two Gibbs distributions $\mu(\cdot \mid v \leftarrow a)$ and $\mu(\cdot \mid v \leftarrow b)$, we pick a free vertex $u \in S_G(v, \ell)$ and consider the maximal coupling of the marginals $\mu(u \leftarrow \cdot \mid v \leftarrow a), \mu(u \leftarrow \cdot \mid v \leftarrow b)$ at u . If u is successfully coupled to a common color c , we proceed to couple $\mu(\cdot \mid v \leftarrow a, u \leftarrow c)$ and $\mu(\cdot \mid v \leftarrow b, u \leftarrow c)$. Otherwise, when u is assigned the disagreement c and d respectively, we couple $\mu(\cdot \mid v \leftarrow a, u \leftarrow c)$ and $\mu(\cdot \mid v \leftarrow b, u \leftarrow d)$ using the path coupling:

- recursively couple $\mu(\cdot \mid v \leftarrow a, u \leftarrow c)$ and $\mu(\cdot \mid v \leftarrow a, u \leftarrow d)$;
- recursively couple $\mu(\cdot \mid v \leftarrow a, u \leftarrow d)$ and $\mu(\cdot \mid v \leftarrow b, u \leftarrow d)$;
- concatenate two couples by path coupling.

Let $\mathcal{W}(\Phi, v)$ be the maximum Wasserstein-1 distance between the two Gibbs distributions obtained by pinning v to two different feasible colors in Φ . Moreover, define

$$\mathcal{W}(G, q) := \max_{\Phi \in \mathcal{L}(G, q), v \in V_\Phi} \mathcal{W}(\Phi, v).$$

The following recursive upper bound for $\mathcal{W}(\Phi, v)$ comes directly from the coupling.

Lemma 4.5. *Under the hypothesis of Corollary 4.4, it holds that*

$$\mathcal{W}(\Phi, v) \leq (D^R + 1) + e^{-\kappa R} H_{D^R} \max_{\Phi', u} \mathcal{W}(\Phi', u)$$

where Φ' is the child instance of Φ obtained by pinning v and possibly some vertices at the sphere $S_{G_\Phi}(v, R)$, and u is an unpinned vertex in $S_{G_\Phi}(v, R)$.

Proof. Write $S = S_{G_\Phi}(v, R)$ and $m = |S| \leq D^R$. Fix σ as a feasible partial assignment on $P \subseteq S$ and a feasible prescription $v \leftarrow c$. Denote by $\Phi^{\sigma, c}$ the list-coloring instance from Φ under the pinning $\sigma \wedge \{v \leftarrow c\}$ and let $U^{\sigma, c} := V_\Phi \setminus (P \cup \{v\})$ be the free vertices of $\Phi^{\sigma, c}$. At first it is clear to see $\Phi^{\sigma, c} \in \mathcal{L}(G, q)$ for every $P \subseteq S$, every feasible partial assignment σ on P and every feasible color c at v .

Set the quantity

$$\mathcal{W}_M := \max_{\substack{P \subseteq S, \text{ feasible } \sigma \text{ on } P, \\ \text{feasible color } v \leftarrow c, u \in S \setminus P}} \mathcal{W}(\Phi^{\sigma, c}, u).$$

It suffices to show $\mathcal{W}(\Phi, v) \leq (D^R + 1) + e^{-\kappa R} H_m \cdot \mathcal{W}_M$. For every $\ell = 0, \dots, m$, define

$$\mathcal{D}_\ell := \max_{\substack{P \subseteq S: |S \setminus P| = \ell \\ \text{feasible } \sigma \text{ on } P, v \leftarrow a \text{ and } v \leftarrow b}} \mathcal{W}_1(\mu_\Phi(\cdot \mid \sigma, v \leftarrow a), \mu_\Phi(\cdot \mid \sigma, v \leftarrow b)).$$

Then $\mathcal{W}(\Phi, v) = \mathcal{D}_m$.

Now we show by induction that

$$\mathcal{D}_\ell \leq \begin{cases} D^R + 1, & \ell = 0 \\ \mathcal{D}_{\ell-1} + \frac{e^{-\kappa R}}{\ell} \mathcal{W}_M, & \ell = 1, \dots, m \end{cases}$$

and thus it concludes $\mathcal{W}(\Phi, v) = \mathcal{D}_m \leq (D^R + 1) + e^{-\kappa R} H_m \cdot \mathcal{W}_M$.

Base case $\ell = 0$. It is clear that the disagreement can only exist in $B_{G_\Phi}(v, R - 1)$. Hence $\mathcal{D}_0 \leq \sum_{j=0}^{R-1} D^j \leq D^R + 1$.

The intermediate case $\ell = 1, \dots, m$. Fix a sphere subset $P \subseteq S$ with $|S \setminus P| = \ell$, a feasible pinning σ on P and an arbitrary free vertex $u \in S \setminus P$. By the one-disagreement coupling in [CLMM26], it holds that for every feasible colors a, b at v and c, d at u ,

$$\begin{aligned} \mathcal{W}_1(\mu_\Phi(\cdot \mid \sigma, v \leftarrow a, u \leftarrow c), \mu_\Phi(\cdot \mid \sigma, v \leftarrow b, u \leftarrow d)) &\leq \mathcal{D}_{\ell-1} + \mathbb{1}[c \neq d] \mathcal{W}(\Phi^{\sigma, a}, u) \\ &\leq \mathcal{D}_{\ell-1} + \mathbb{1}[c \neq d] \mathcal{W}_M. \end{aligned}$$

Therefore for every free vertex $u \in S \setminus P$

$$\mathcal{D}_\ell \leq \mathcal{D}_{\ell-1} + \|\mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow a) - \mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow b)\|_{\text{TV}} \cdot \mathcal{W}_M.$$

By Corollary 4.4,

$$\sum_{u \in S \setminus P} \|\mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow a) - \mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow b)\|_{\text{TV}} \leq e^{-\kappa R}.$$

Pick the vertex u to minimize $\|\mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow a) - \mu_\Phi(u \leftarrow \cdot \mid \sigma, v \leftarrow b)\|_{\text{TV}}$, and we have

$$\mathcal{D}_\ell \leq \mathcal{D}_{\ell-1} + \frac{e^{-\kappa R}}{\ell} \mathcal{W}_M.$$

Combining all things together, we conclude that $\mathcal{W}(\Phi, v) \leq (D^R + 1) + e^{-\kappa R} H_m \cdot \mathcal{W}_M$, which leads to the lemma by monotonicity of harmonic numbers. $\square$

4.2. Message passing on marked trees. In this part, we show a message passing process on a local tree with nodes marked as good or bad. The messages would help us analyze the disagreements through the coupling on random graphs.

For a finite tree T rooted at ϱ , each node ω in T is labeled as good or bad and equipped with cost $\text{cost}(\omega) \geq 0$. Let $\text{child}_T(\omega)$ be the child nodes of ω in T and $\text{parent}_T(\omega)$ be the parent node of ω in T with convention $\text{parent}_T(\varrho) = \emptyset$. Let $\text{SubTree}_T(\omega)$ be the subtree of T rooted at ω . For $\alpha \in (0, 1)$ recursively define the message

$$(2) \quad \phi(\alpha; T, \varrho) = \text{cost}(\varrho) + \begin{cases} \alpha \cdot \max_{\omega \in \text{child}_T(\varrho)} \phi(\alpha; \text{SubTree}_T(\omega), \omega), & \varrho \text{ is good} \\ \sum_{\omega \in \text{child}_T(\varrho)} \phi(\alpha; \text{SubTree}_T(\omega), \omega), & \varrho \text{ is bad} \end{cases}$$

where the summation and the maximum over an empty set are defined as zero. For simplicity, for a unrooted tree T , we write $\phi(\alpha; T, \varrho)$ as the message after rooting T at ϱ . The following rerooting lemma records the change of messages.

Lemma 4.6. *For a node ω in T at distance ℓ from ϱ , it holds that*

$$\phi(\alpha; T - \varrho, \omega) \leq \ell \alpha^{-\ell} \phi(\alpha; T, \varrho)$$

where $\phi(\alpha; T - \varrho, \omega)$ only needs to pass messages on the component of $T - \varrho$ containing ω after switching the root to ω , with inherited marks and costs.

Proof. Assume that the unique path from ϱ to ω is

$$\varrho = \gamma_0, \gamma_1, \dots, \gamma_\ell = \omega.$$

Remove all edges of this path. Let T_i be the remaining component contains γ_i for $i = 0, \dots, \ell$.

We firstly observe that

$$\phi(\alpha; T_i, \gamma_i) \leq \phi(\alpha; \text{SubTree}_T(\gamma_i), \gamma_i).$$

In fact, T_i is a subtree of $\text{SubTree}_T(\gamma_i)$ and the upper bound is direct. Furthermore,

$$\phi(\alpha; \text{SubTree}_T(\gamma_i), \gamma_i) \leq \alpha^{-1} \phi(\alpha; \text{SubTree}_T(\gamma_{i-1}), \gamma_{i-1})$$

by recursion of the message ϕ . Therefore for every $i = 0, \dots, \ell$,

$$\phi(\alpha; T_i, \gamma_i) \leq \alpha^{-i} \phi(\alpha; \text{SubTree}_T(\gamma_0), \gamma_0) = \alpha^{-i} \phi(\alpha; T, \varrho).$$

Now for $i = 1, \dots, \ell$, let B_i be the tree consisting of $T_1, \dots, T_i$, joined by edges in path $\gamma_1, \dots, \gamma_i$. Root B_i at γ_i . Then it holds that

$$\phi(\alpha; B_1, \gamma_1) = \phi(\alpha; T_1, \gamma_1), \quad \phi(\alpha; B_\ell, \gamma_\ell) = \phi(\alpha; T - \varrho, \omega).$$

For $i = 2, \dots, \ell$, if γ_i is bad, then we directly have

$$\phi(\alpha; B_i, \gamma_i) = \phi(\alpha; B_{i-1}, \gamma_{i-1}) + \phi(\alpha; T_i, \gamma_i).$$

Otherwise, since $\alpha < 1$, it directly holds that

$$\phi(\alpha; B_i, \gamma_i) \leq \phi(\alpha; B_{i-1}, \gamma_{i-1}) + \phi(\alpha; T_i, \gamma_i)$$

Summing over $i = 2, \dots, \ell$, it holds that

$$\begin{aligned}\phi(\alpha; B_\ell, \gamma_\ell) &\leq \sum_{i=1}^{\ell} \phi(\alpha; T_i, \gamma_i) \\ &\leq \phi(\alpha; T, \varrho) \sum_{i=1}^{\ell} \alpha^{-i} \\ &\leq \ell \alpha^{-\ell} \cdot \phi(\alpha; T, \varrho)\end{aligned}$$

which concludes this rerooting lemma. $\square$

4.3. Coupling independence by local messages. The next step is to bound the coupling independence by local messages.

For a subgraph J of G such that it is tree or unicyclic, we construct a tree $\mathcal{T} = \mathcal{T}(J)$ as:

- if J is a tree itself, let $\mathcal{T}$ as a copy of J and $\text{node}(\cdot) : V(J) \rightarrow V(\mathcal{T})$ be the mapping such that it maps every vertex u in J to its corresponding node in $\mathcal{T}$;
- otherwise J is unicyclic. Let C be the unique cycle in J . We contract all vertices in C into one cycle node and retain other vertices as single nodes. For every $(u, v) \in E(J)$, there is an edge between $\text{node}(u)$ and $\text{node}(v)$.

Furthermore, we mark nodes in $\mathcal{T}$ as bad or good in the following way.

- At first, we mark $\omega \in \mathcal{T}$ as bad if it is a cycle node or its corresponding vertex is in V^{HD} .
- For every currently unmarked node ω , if there exists a node marked as bad in the first step within distance r from ω , mark ω as bad.
- Finally mark all remaining nodes as good.

Lemma 4.7. *Let J be an induced subgraph of G which is tree or unicyclic. Furthermore, fix the boundary $\Sigma \subseteq V(J)$ as:*

$$\Sigma := \{u \in V(J) : \text{dist}_G(u, V(G) \setminus V(J)) \leq r + R + 3\}.$$

We consider $\mathcal{T} = \mathcal{T}(J)$ with marks on nodes. Define the cost on $\mathcal{T}$ as:

- $\text{cost}(\omega) = 2(D^R + 1) \cdot |C \setminus \Sigma| + \mathcal{W}(G, q) \cdot |C \cap \Sigma|$ if ω corresponds to the unique cycle C ;
- $\text{cost}(\omega) = 2(D^R + 1)$ if ω corresponds to a vertex not in Σ ;
- $\text{cost}(\omega) = \mathcal{W}(G, q)$ if ω corresponds to a vertex in Σ .

Then it holds that

$$\mathcal{W}(\Phi, v) \leq \phi(\alpha; \mathcal{T}, \text{node}(v))$$

for every $\alpha \in [\rho, 1)$, feasible list-coloring instance $\Phi \in \mathcal{L}(G, q)$ and $v \in V_\Phi \cap V(J)$.

Proof. Fix a feasible list-coloring instance $\Phi \in \mathcal{L}(G, q)$. Let $\mathcal{T}[\Phi]$ be the forest generated from $\mathcal{T}$ by deleting all nodes with no vertices in V_Φ . Set $\mathcal{T}[\Phi](v)$ be the tree in $\mathcal{T}[\Phi]$ containing the corresponding node of v . We take d sufficiently large that $q - D \geq 2$. We prove by induction on $|V_\Phi|$ that

$$(3) \quad \mathcal{W}(\Phi, v) \leq \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)).$$

The boundary case: $v \in \Sigma$. By the definition of the costs, including the case where $\text{node}(v)$ is a cycle node,

$$\mathcal{W}(\Phi, v) \leq \mathcal{W}(G, q) \leq \text{cost}(\text{node}(v)) \leq \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)).$$

The first interior case: $v \notin \Sigma$ and $\text{node}(v)$ is bad. Let B be the maximal connected component of bad nodes containing $\text{node}(v)$ in $\mathcal{T}[\Phi]$, and define

$$U := \{x \in V(J) \cap V_\Phi : \text{node}(x) \in B\}, \quad W := V_\Phi \setminus U.$$

Suppose first that $U \cap \Sigma \neq \emptyset$, and choose $x \in U \cap \Sigma$. Every node on the path from $\text{node}(v)$ to $\text{node}(x)$ is bad. Consequently, the sum rule in the message recursion gives

$$\phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) \geq \text{cost}(\text{node}(x)) \geq \mathcal{W}(G, q) \geq \mathcal{W}(\Phi, v).$$

Thus we may assume that $U \cap \Sigma = \emptyset$. Every vertex of U has all its neighbors in J . Let $\mathcal{T}_1, \dots, \mathcal{T}_\ell$ be the components of $\mathcal{T}[\Phi] - B$ adjacent to B . Since $\mathcal{T}[\Phi]$ is a forest and a retained cycle node is bad, the edges of G_Φ between U and W can be indexed as

$$(u_1, v_1), \dots, (u_\ell, v_\ell), \quad u_j \in U, \quad v_j \in W, \quad \text{node}(v_j) \in V(\mathcal{T}_j).$$

Each $\mathcal{T}_j$ has exactly one such edge. Moreover, the vertices $v_1, \dots, v_\ell$ are distinct, their corresponding nodes are good, and $\deg_G(v_j) \leq D$. Expanding the sum rule throughout B , we obtain

$$(4) \quad \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) = \sum_{\omega \in B} \text{cost}(\omega) + \sum_{j=1}^{\ell} \phi(\rho; \mathcal{T}_j, \text{node}(v_j)).$$

Since $U \cap \Sigma = \emptyset$, the first sum is at least $|U|$. Fix two feasible colors at v , and couple the corresponding marginals on U arbitrarily. For a pair of resulting assignments σ, τ on U , we now couple the conditional distributions on the *entire* remaining vertex set W . For $j = 0, \dots, \ell$, let $\Phi^{(j)}$ be the list-coloring instance on $G[W]$ with lists

$$L^{(j)}(x) = \begin{cases} L_\Phi(v_i) \setminus \{\tau(u_i)\}, & x = v_i, \ i \leq j, \\ L_\Phi(v_i) \setminus \{\sigma(u_i)\}, & x = v_i, \ i > j, \\ L_\Phi(x), & x \notin \{v_1, \dots, v_\ell\}. \end{cases}$$

Write $\nu^{(j)} = \mu_{\Phi^{(j)}}$. Then $\nu^{(0)}$ and $\nu^{(\ell)}$ are precisely the conditional distributions on W given σ and τ , respectively.

We verify that these intermediate instances are feasible. The instance $\Phi^{(0)}$ is feasible by construction. At step j , only the list at v_j changes. If a coloring of $\Phi^{(j-1)}$ is not valid for $\Phi^{(j)}$, recolor v_j to a color in

$$L_\Phi(v_j) \setminus (\{\tau(u_j)\} \cup \{\xi(y) : y \in N_{G[W]}(v_j)\}),$$

where ξ denotes the current coloring. Such a color exists because

$$|L_\Phi(v_j)| \geq \deg_{G_\Phi}(v_j) + q - \deg_G(v_j) = \deg_{G[W]}(v_j) + 1 + q - \deg_G(v_j) \geq \deg_{G[W]}(v_j) + 3.$$

Hence every $\Phi^{(j)}$ is feasible. Furthermore, for every $x \in W$,

$$|L^{(j)}(x)| - \deg_{G[W]}(x) \geq |L_\Phi(x)| - \deg_{G_\Phi}(x) \geq q - \deg_G(x),$$

so $\Phi^{(j)} \in \mathcal{L}(G, q)$.

For $j = 1, \dots, \ell$, define $\widehat{\Phi}_j$ on $G[W]$ by using the common lists of $\Phi^{(j-1)}$ and $\Phi^{(j)}$ at all vertices other than v_j , and restoring the list $L_\Phi(v_j)$ at v_j . This is a feasible instance in $\mathcal{L}(G, q)$, and

$$\nu^{(j-1)} = \mu_{\widehat{\Phi}_j}(\cdot \mid v_j \neq \sigma(u_j)), \quad \nu^{(j)} = \mu_{\widehat{\Phi}_j}(\cdot \mid v_j \neq \tau(u_j)).$$

Both distributions are mixtures of the same family

$$\left\{ \mu_{\widehat{\Phi}_j}(\cdot \mid v_j = c) : c \text{ is feasible at } v_j \right\}.$$

Coupling the mixing colors and then optimally coupling the corresponding conditional distributions yields

$$\mathcal{W}_1(\nu^{(j-1)}, \nu^{(j)}) \leq \mathcal{W}(\widehat{\Phi}_j, v_j).$$

Since $V_{\widehat{\Phi}_j} = W$ and $v \in U$, these instances have fewer free vertices than Φ . Moreover,

$$\mathcal{T}[\widehat{\Phi}_j] = \mathcal{T}[\Phi] - B, \quad \mathcal{T}[\widehat{\Phi}_j](v_j) = \mathcal{T}_j.$$

The induction hypothesis and the triangle inequality give

$$\mathcal{W}_1(\nu^{(0)}, \nu^{(\ell)}) \leq \sum_{j=1}^{\ell} \mathcal{W}(\widehat{\Phi}_j, v_j) \leq \sum_{j=1}^{\ell} \phi(\rho; \mathcal{T}_j, \text{node}(v_j)).$$

Combining this conditional coupling on W with the coupling on U , and then maximizing over the two feasible colors at v , we obtain

$$\mathcal{W}(\Phi, v) \leq |U| + \sum_{j=1}^{\ell} \phi(\rho; \mathcal{T}_j, \text{node}(v_j)) \leq \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v))$$

where the last inequality follows from (4).

The second interior case: $v \notin \Sigma$ and $\text{node}(v)$ is marked as good. By the marking process, $B_G(v, r) \subseteq J$ and $B_G(v, r) \cap V^{\text{HD}} = \emptyset$. Lemma 4.5 provides that

$$\mathcal{W}(\Phi, v) \leq (D^R + 1) + e^{-\kappa R} H_{D^R} \cdot \max_{\Phi', u} \mathcal{W}(\Phi', u).$$

Here each child instance Φ' is defined on a subset $V_{\Phi'} \subseteq V_{\Phi} \setminus \{v\}$ and has a vertex $u \in V_{\Phi'} \cap B_{G_{\Phi}}(v, R)$. Furthermore, the R -length path from v to u completely lies in the intersection of J and G_{Φ} and therefore there is a path from $\text{node}(v)$ to $\text{node}(u)$ of length at most R in $\mathcal{T}[\Phi]$. By the induction hypothesis and Lemma 4.6,

$$\begin{aligned} \mathcal{W}(\Phi', u) &\leq \phi(\rho; \mathcal{T}[\Phi'](u), u) \\ &\leq \phi(\rho; \mathcal{T}[\Phi](v) - \text{node}(v), \text{node}(u)) \\ &\leq \ell \rho^{-\ell} \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)). \\ &\leq R \rho^{-R} \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) \end{aligned}$$

where the last inequality holds for $\ell \leq R$ and $\rho < 1$. Recall Corollary 4.4, and we obtain that

$$\begin{aligned} \mathcal{W}(\Phi, v) &\leq (D^R + 1) + e^{-\kappa R} H_{D^R} \cdot R \rho^{-R} \cdot \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) \\ &\leq (D^R + 1) + \frac{1}{4} \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) \\ &\leq \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)) \end{aligned}$$

where the last inequality holds for

$$(D^R + 1) \leq \frac{1}{2} \text{cost}(\text{node}(v)) \leq \frac{1}{2} \phi(\rho; \mathcal{T}[\Phi](v), \text{node}(v)).$$

Putting all things together, we conclude the lemma with monotonicity of ϕ on $\alpha \geq \rho$. $\square$

4.3.1. *Further decomposition of messages.* To bound $\phi(\alpha; \mathcal{T}, \text{node}(v))$ in Lemma 4.7, it is convenient to introduce the following decomposition.

Let $Y(\alpha; \mathcal{T}, \text{node}(v)), Z(\alpha; \mathcal{T}, \text{node}(v))$ be the messages on $\mathcal{T}$ as (2) with retained marks and the costs defined as

$$\begin{aligned} \text{cost}_Y(\omega) &= \begin{cases} 2(D^R + 1) \cdot |C|, & \omega \text{ corresponds to the unique cycle } C \\ 2(D^R + 1), & \omega \text{ corresponds to a single vertex } v \end{cases}, \\ \text{cost}_Z(\omega) &= \begin{cases} |C \cap \Sigma|, & \omega \text{ corresponds to the unique cycle } C \\ \mathbb{1}[v \in \Sigma], & \omega \text{ corresponds to a single vertex } v \end{cases}. \end{aligned}$$

Proposition 4.8. *The following properties hold.*

- $\phi(\alpha; \mathcal{T}, \text{node}(v)) \leq Y(\alpha; \mathcal{T}, \text{node}(v)) + \mathcal{W}(G, q)Z(\alpha; \mathcal{T}, \text{node}(v)).$
- $Z(\alpha; \mathcal{T}, \text{node}(v)) \leq \frac{Y(\alpha; \mathcal{T}, \text{node}(v))}{2(D^R + 1)}.$
- Assume that every path from $\text{node}(v)$ to a node of positive cost_Z contains at least ℓ good nodes other than its endpoint. Then for every $0 < \alpha_0 \leq \alpha_1 < 1$,

$$Z(\alpha_0; \mathcal{T}, \text{node}(v)) \leq \left(\frac{\alpha_0}{\alpha_1} \right)^\ell Z(\alpha_1; \mathcal{T}, \text{node}(v)).$$

Proof. Set $A := 2(D^R + 1)$ and $W := \mathcal{W}(G, q)$. The update in (2) is monotone, positively homogeneous, and subadditive, where subadditivity at a good node follows from

$$\max_i (x_i + y_i) \leq \max_i x_i + \max_i y_i.$$

At every node, the three costs satisfy

$$\text{cost} \leq \text{cost}_Y + W \text{cost}_Z, \quad \text{cost}_Z \leq A^{-1} \text{cost}_Y.$$

For a cycle node these inequalities reduce to $|C \cap \Sigma| \leq |C|$; for a single-vertex node they are immediate. Induction on the height of the rooted subtree, using the three properties of the update, proves the first two claims.

For the last claim, put $\theta := \alpha_0/\alpha_1$. For a node ω , let $g(\omega)$ be the minimum number of good nonterminal nodes on a path from ω to a node of positive cost_Z in its descendant subtree. If no such node exists, its Z -message is zero. We claim, again by induction on subtree height, that

$$(5) \quad Z(\alpha_0; \mathcal{T}_\omega, \omega) \leq \theta^{g(\omega)} Z(\alpha_1; \mathcal{T}_\omega, \omega).$$

If $\text{cost}_Z(\omega) > 0$, then $g(\omega) = 0$, and the claim follows from monotonicity in α . Otherwise, writing g_i for the corresponding values at the children and applying the induction hypothesis gives

$$\begin{aligned} Z(\alpha_0; \mathcal{T}_\omega, \omega) &\leq \theta^{\min_i g_i} Z(\alpha_1; \mathcal{T}_\omega, \omega), & \omega \text{ bad,} \\ Z(\alpha_0; \mathcal{T}_\omega, \omega) &\leq \theta^{1+\min_i g_i} Z(\alpha_1; \mathcal{T}_\omega, \omega), & \omega \text{ good.} \end{aligned}$$

These exponents are precisely $g(\omega)$, proving (5). The assumption gives $g(\text{node}(v)) \geq \ell$, and the third claim follows since $0 < \theta \leq 1$. $\square$

4.4. **Estimation to local messages.** On the event $\mathcal{E}_{5L}^{\text{local}}$, for every vertex v , we pick J_v as:

$$J_v = \begin{cases} B_G(v, 2L), & \text{if } B_G(v, 2L) \text{ is a tree} \\ B_G(v, 4L), & \text{if } B_G(v, 2L) \text{ is unicyclic} \end{cases}.$$

Let $\mathcal{T}_v$ be the tree induced by J_v as in Lemma 4.7. Now we give the following upper bound to $Y(\alpha; \mathcal{T}_v, \text{node}(v))$.

Lemma 4.9. Fix $\alpha \in (0, 1)$. For d sufficiently large, depending on η and α , there exists a constant $C = C(\eta, d, q, \alpha) > 0$ such that, with probability at least $1 - o_n(1)$, the event $\mathcal{E}_{5L}^{\text{local}}$ holds and

$$Y(\alpha; \mathcal{T}_v, \text{node}(v)) \leq C \ln n \quad \text{for every } v \in V.$$

To prove Lemma 4.9, we consider the message passing on a Galton–Watson tree. Let $(\mathcal{T}^{\text{GW}}, \varrho)$ be a Galton–Watson tree rooted at ϱ such that for every node $\omega \in \mathcal{T}^{\text{GW}}$, it independently has $K_\omega \sim \text{Poisson}(d+1)$ children. Equip $\mathcal{T}^{\text{GW}}$ with the depth $\text{depth}(\omega) := \text{dist}_{\mathcal{T}^{\text{GW}}}(\varrho, \omega)$. Let $\mathcal{H}$ be the collection of nodes of more than $D-3$ children, i.e.,

$$\mathcal{H} := \{\omega \in \mathcal{T}^{\text{GW}} \mid K_\omega > D-3\}$$

and choose a parameter h as

$$h := \left\lceil \max \left\{ 3r, \frac{4 \ln 4}{\ln(1/\alpha)} \right\} \right\rceil.$$

We mark nodes in $\mathcal{T}^{\text{GW}}$ as following two ways:

$$\mathcal{B} := \{\omega \in \mathcal{T}^{\text{GW}} \mid \text{dist}_{\mathcal{T}^{\text{GW}}}(\omega, \mathcal{H}) \leq r\}, \quad \mathcal{B}^+ := \mathcal{B} \cup \{\omega \in \mathcal{T}^{\text{GW}} \mid \text{depth}(\omega) < 2h\}.$$

We define the message on $\mathcal{T}^{\text{GW}}$ as the limit of messages on finite depth. Formally, for $\ell \geq 0$ and a subset of nodes S representing bad nodes, for every $\text{depth}(\omega) \leq \ell$,

$$\phi_{\ell, S}(\alpha; \omega) = 2(D^R + 1) + \begin{cases} \alpha \cdot \max_{\gamma \in \text{child}_{\mathcal{T}^{\text{GW}}}(\omega)} \phi_{\ell, S}(\alpha; \gamma) & \omega \notin S \\ \sum_{\gamma \in \text{child}_{\mathcal{T}^{\text{GW}}}(\omega)} \phi_{\ell, S}(\alpha; \gamma) & \omega \in S \end{cases}$$

with convention that the maximum over an empty set is zero. Define

$$\mathcal{X} := \lim_{\ell \rightarrow \infty} \phi_{\ell, \mathcal{B}}(\alpha), \quad \mathcal{Y} := \lim_{\ell \rightarrow \infty} \phi_{\ell, \mathcal{B}^+}(\alpha).$$

We first clarify here that $\mathcal{X}$ and $\mathcal{Y}$ are well-defined and finite almost surely for all sufficiently large $d > 1$.

Proposition 4.10. It holds that $\mathcal{X}$ and $\mathcal{Y}$ exist. Furthermore, for sufficiently large $d > 1$, there exists some $t_0 > 0$ such that

$$\mathbb{E}[e^{t_0 \mathcal{X}}] < \infty, \quad \mathbb{E}[e^{t_0 \mathcal{Y}}] < \infty.$$

The proof of Proposition 4.10 is provided in Appendix B. Here we show how to upper bound local messages by Proposition 4.10.

Proof of Lemma 4.9. Fix $\alpha \in (0, 1)$ and take d sufficiently large for Proposition 4.10 to apply. Write

$$a_0 := 2(D^R + 1), \quad \lambda := d + 1, \quad L_0 := 4L + 2r + 3, \quad \mathcal{E} := \mathcal{E}_{5L}^{\text{local}}.$$

Since r is fixed as $n \rightarrow \infty$, we have $L_0 \leq 5L$ for all sufficiently large n . Consequently, on $\mathcal{E}$, every ball $B_G(v, L_0)$ is a tree or unicyclic. Therefore the prescribed choice of J_v is well-defined. On this event write

$$Y_v := Y(\alpha; \mathcal{T}_v, \text{node}(v))$$

and for notational convenience, set $Y_v = 0$ otherwise.

By Proposition 4.10, there exists $t_0 > 0$ such that $\mathbb{E}[e^{t_0 \mathcal{X}}]$ and $\mathbb{E}[e^{t_0 \mathcal{Y}}]$ are finite. Choose $t \in (0, t_0]$ sufficiently small that

$$M_X := \mathbb{E}[e^{t \mathcal{X}}] < \infty, \quad M_Y := \mathbb{E}[e^{t \mathcal{Y}}] < \infty, \quad ta_0 + \lambda(M_Y - 1) \leq 1.$$

Such a choice is possible because, by dominated convergence, $M_Y \rightarrow 1$ as $t \downarrow 0$.

To compare local messages on the original graph with $\mathcal{X}$ and $\mathcal{Y}$, we first describe the following exploration coupling. Reveal edges adaptively without conditioning on the absence of future collisions. Conditional on the revealed history, the edges from the current vertex to undiscovered

vertices that have not previously been queried are independent Bernoulli variables with parameter d/n . Thus their number has distribution $\text{Binomial}(N, d/n)$ for some $N \leq n$, and

$$\text{Binomial}(N, d/n) \preceq \text{Poisson}(-N \ln(1 - d/n)) \preceq \text{Poisson}(\lambda)$$

for all sufficiently large n .

At successive exploration steps, use fresh independent randomness to realize this domination. Complete all unused offspring slots independently. The resulting auxiliary offspring variables are independent $\text{Poisson}(\lambda)$ variables. If the exploration reveals a surplus edge not prescribed in advance, stop the comparison and complete the auxiliary trees independently. We use the comparison only on events on which no such surplus edge occurs. We mark that the exploration is carried out far enough to determine the relevant degrees and marks since the exploration is possible inside $B_G(v, L_0)$ because $J_v \subseteq B_G(v, 4L)$ and $4L + r + 1 < L_0$.

The first case: tree balls. Fix $v \in V$ and let

$$\mathcal{A}_v^T := \{B_G(v, L_0) \text{ is a tree}\}.$$

Apply the preceding exploration coupling from v with no prescribed surplus edge. On $\mathcal{A}_v^T$, the exploration gives a rooted embedding of J_v into the Galton–Watson tree. Every relevant non-root vertex has one parent edge and all its other incident edges are counted among its children. Hence a vertex of degree greater than D is mapped to a node whose auxiliary offspring count is greater than $D - 3$. The embedding preserves the tree paths, so every bad node of $\mathcal{T}_v$ is mapped to a node of $\mathcal{B}$.

Adding branches and changing a good rule to a bad rule can only increase a message with non-negative costs. Therefore, under this coupling,

$$Y_v \leq \mathcal{X} \quad \text{on } \mathcal{A}_v^T.$$

The comparison here is event-wise. In particular,

$$\mathbb{E} [\mathbb{1} [\mathcal{A}_v^T e^{tY_v}]] \leq M_X.$$

By Markov inequality, for every $x \geq 0$,

$$(6) \quad \mathbb{P} (\mathcal{A}_v^T \cap \{Y_v > x\}) \leq M_X e^{-tx}.$$

The second case: unicyclic balls. A rooted core (v, H) consists of a simple cycle together with a simple path from v to that cycle, meeting the cycle only at the endpoint of the path. The path is allowed to have length zero. In particular, $|V(H)| = |E(H)|$. For a fixed deterministic rooted core (v, H) , define the event $\mathcal{A}_{v,H}$ as the joint event of

$$\{B_G(v, L_0) \text{ is unicyclic}\} \wedge \{H \text{ is its unique cycle together with the path from } v \text{ to that cycle}\}.$$

Obviously, $\mathcal{A}_{v,H} \subseteq \{H \subseteq G\}$.

Let $k = |V(H)|$. Conditional on the event $H \subseteq G$, all edges outside $E(H)$ remain independent Bernoulli variables with parameter d/n . Explore outwards from the prescribed core. Give each core vertex its distance from v within H as its initial distance label, and explore in increasing order of these labels and the subsequently discovered distances, up to distance L_0 . Stop the comparison if an additional surplus edge is discovered. On $\mathcal{A}_{v,H}$, no such additional surplus edge occurs in the explored ball, and the discovered branches are trees attached to H . Force the vertices of $H \cap J_v$ to be bad and contract this connected root set. If the cycle is present in J_v , use its already contracted cycle node. Increase the cost of the resulting root to ka_0 . These operations cannot decrease the message: replacing a good rule by the sum rule increases it, and contracting connected bad nodes adds their costs and still retains their outgoing branches.

The total number of branches leaving the core is dominated by

$$K_k \sim \text{Poisson}(k\lambda),$$

obtained by summing k independent $\text{Poisson}(\lambda)$ variables. All descendant auxiliary trees can be chosen independently of one another and of K_k . Within each branch, high-degree marks are dominated by the high-offspring marks. Any additional bad mark caused by the contracted core or by a high-degree vertex outside that branch can affect only vertices within distance r of the core. These vertices are covered by the extra bad levels in the definition of $\mathcal{Y}$, since $h \geq 3r$ and $\mathcal{B}^+$ contains every node of depth less than $2h$.

Consequently, conditional on $H \subseteq G$, the exploration admits a coupling with

$$Z_k := ka_0 + \sum_{j=1}^{K_k} \mathcal{Y}_j$$

where the $\mathcal{Y}_j$ are independent copies of $\mathcal{Y}$ independent of K_k , such that

$$Y_v \leq Z_k \quad \text{on } \mathcal{A}_{v,H}.$$

The exponential moment of Z_k satisfies

$$\mathbb{E} [e^{tZ_k}] = \exp(kt a_0 + k\lambda(M_Y - 1)) \leq e^k.$$

Hence

$$\mathbb{P}(\mathcal{A}_{v,H} \cap \{Y_v > x\} \mid H \subseteq G) \leq e^{k-tx}.$$

Since $\mathcal{A}_{v,H} \subseteq \{H \subseteq G\}$ and H has exactly k edges, we obtain the unconditional joint-event bound

$$(7) \quad \mathbb{P}(\mathcal{A}_{v,H} \cap \{Y_v > x\}) \leq \left(\frac{d}{n}\right)^k e^{k-tx}.$$

Now we count the number of possible cores. If $B_G(v, L_0)$ is unicyclic, its unique cycle has length at most $2L_0 + 1$, and the path from v to that cycle has length at most L_0 . Thus its rooted core satisfies

$$3 \leq k \leq k_{\max} := 3L_0 + 1.$$

Let $\mathfrak{H}_k$ be the collection of all deterministic rooted cores (v, H) on k vertices in $[n]$. To encode such a core, choose an ordered list of k distinct vertices, starting with v , and choose the position at which the path ends and the cycle begins. The cycle is then closed by joining the last vertex to that position. This gives the upper bound $|\mathfrak{H}_k| \leq kn^k$.

On $\mathcal{E}$, every vertex belongs to one of the tree or unicyclic cases considered above. Applying the union bound to intersections with $\mathcal{E}$, and then using (6) and (7), gives

$$\begin{aligned} \mathbb{P}(\mathcal{E} \cap \{\exists v \in V : Y_v > x\}) &\leq \sum_{v \in V} \mathbb{P}(\mathcal{A}_v^T \cap \{Y_v > x\}) + \sum_{k=3}^{k_{\max}} \sum_{(v,H) \in \mathfrak{H}_k} \mathbb{P}(\mathcal{A}_{v,H} \cap \{Y_v > x\}) \\ &\leq nM_X e^{-tx} + \sum_{k=3}^{k_{\max}} kn^k \left(\frac{d}{n}\right)^k e^{k-tx} \\ &= e^{-tx} \left(nM_X + \sum_{k=3}^{k_{\max}} k(ed)^k \right). \end{aligned}$$

Recall that

$$L = \left\lfloor \frac{\ln n}{1000 \ln(2ed)} \right\rfloor.$$

Since r is fixed as $n \rightarrow \infty$, we have $k_{\max} = 12L + O_{\eta,d}(1)$ and therefore

$$\sum_{k=3}^{k_{\max}} k(ed)^k \leq k_{\max}^2 (ed)^{k_{\max}} = n^{\frac{12 \ln(ed)}{1000 \ln(2ed)} + o_n(1)} = o(n).$$

Taking $x = (4/t) \ln n$, we conclude that

$$\mathbb{P} \left(\mathcal{E} \cap \left\{ \exists v \in V : Y_v > \frac{4}{t} \ln n \right\} \right) = O(n^{-3}).$$

Finally, by Lemma 3.1, $\mathbb{P}(\overline{\mathcal{E}}) = o_n(1)$. Thus

$$\mathbb{P} \left(\mathcal{E} \cap \left\{ Y_v \leq \frac{4}{t} \ln n \text{ for every } v \in V \right\} \right) \geq 1 - \mathbb{P}(\overline{\mathcal{E}}) - O(n^{-3}) = 1 - o_n(1)$$

which concludes the lemma with $C = 4/t$. $\square$

4.5. Logarithmic bound for coupling independence. In this section, we complete the proof of Theorem 1.4

Lemma 4.11. *Assume $\mathcal{E}_{5L}^{\text{local}} \cap \mathcal{E}_{\lfloor L/4 \rfloor, r}^{\text{good}}$. For all sufficiently large n , every path in $\mathcal{T}_v$ from $\text{node}(v)$ to a node of positive cost_Z contains at least $\lfloor L/8 \rfloor$ good nodes other than its endpoint.*

Proof. Set $L_0 = \lfloor L/4 \rfloor$ and $c_0 = r + R + 3$. If $J_v = B_G(v, 2L)$, then every $u \in \Sigma$ satisfies

$$\text{dist}_G(v, u) \geq 2L + 1 - c_0.$$

If $J_v = B_G(v, 4L)$, its unique cycle C is contained in $B_G(v, 2L)$, and hence

$$\text{dist}_G(u, C) \geq 2L + 1 - c_0 \quad \text{for every } u \in \Sigma.$$

In particular, $C \cap \Sigma = \emptyset$, so every node of positive cost_Z corresponds to a single vertex $u \in \Sigma$. Thus, for sufficiently large n , the path from $\text{node}(v)$ to $\text{node}(u)$ contains an internal subpath P on L_0 vertices which is disjoint from the r -neighborhood of the cycle node and does not contain the terminal node.

The subpath P corresponds to a simple path in G . By $\mathcal{E}_{L_0, r}^{\text{good}}$, at least $L_0/2$ of its vertices lie outside $B_G(V^{\text{HD}}, r)$. These vertices are also at distance greater than r from the cycle and therefore are good. Since

$$\lceil L_0/2 \rceil \geq \lfloor L/8 \rfloor,$$

we conclude the lemma. $\square$

Now we are ready to prove Theorem 1.4.

Proof of Theorem 1.4. Pick L as in Section 3 and κ, ρ, R and r as Corollary 4.4. We assume $\mathcal{E}_{5L}^{\text{local}}$ and $\mathcal{E}_{\lfloor L/4 \rfloor, r}^{\text{good}}$. The joint probability is at least $1 - o_n(1)$ by Lemmas 3.1 and 3.3.

For every $\Phi \in \mathcal{L}(G, q)$ and v , by Lemma 4.11 and Lemma 4.7 and Proposition 4.8, applied with coefficients ρ and $\rho^{1/2}$, it holds that

$$\mathcal{W}(\Phi, v) \leq Y(\rho^{1/2}; \mathcal{T}, \text{node}(v)) + \frac{\mathcal{W}(G, q)}{2(D^R + 1)} \rho^{\lfloor L/8 \rfloor / 2} Y(\rho^{1/2}; \mathcal{T}, \text{node}(v)).$$

By Lemma 4.9, with probability at least $1 - o_n(1)$, $Y(\rho^{1/2}; \mathcal{T}, \text{node}(v)) \leq C_{\eta, d, q} \ln n$ for some constant $C_{\eta, d, q} > 0$. Then for sufficiently large n ,

$$\mathcal{W}(\Phi, v) \leq C_{\eta, d, q} \ln n + \varepsilon_n \mathcal{W}(G, q)$$

where $\varepsilon_n = O(\rho^{\lfloor L/8 \rfloor / 2} \ln n) \rightarrow 0$ as $n \rightarrow \infty$. Take maximum on the left side to obtain that

$$\mathcal{W}(G, q) \leq C_{\eta, d, q} \ln n + \varepsilon_n \mathcal{W}(G, q)$$

meaning that $\mathcal{W}(G, q) \leq C' \ln n$ for some constant $C' > 0$ depending only on η, d and q . Therefore we conclude the theorem. $\square$

Proof of Corollary 1.5. Fix a feasible pinning τ and a feasible source pair $v \leftarrow a$. For every $w \neq v$ and color c ,

$$\Psi_\mu^\tau(v \leftarrow a, w \leftarrow c) = \sum_b \mu^\tau(v \leftarrow b)(\mu^\tau(w \leftarrow c \mid v \leftarrow a) - \mu^\tau(w \leftarrow c \mid v \leftarrow b)).$$

Therefore,

$$\begin{aligned} \sum_{w \neq v} \sum_c |\Psi_\mu^\tau(v \leftarrow a, w \leftarrow c)| &\leq 2 \sum_b \mu^\tau(v \leftarrow b) \sum_{w \neq v} \left\| \mu_w^{\tau, v \leftarrow a} - \mu_w^{\tau, v \leftarrow b} \right\|_{\text{TV}} \\ &\leq 2 \sum_b \mu^\tau(v \leftarrow b) \mathcal{W}_1(\mu^{\tau, v \leftarrow a}, \mu^{\tau, v \leftarrow b}) \\ &\leq 2 \text{Cl}(\mu). \end{aligned}$$

Hence $\|\Psi_\mu^\tau\|_\infty \leq 2 \text{Cl}(\mu)$, leading to the upper bound that $\lambda_{\max}(\Psi_\mu^\tau) \leq 2 \text{Cl}(\mu)$. The conclusion follows from Theorem 1.4. $\square$

5. RAPID MIXING OF GLAUBER DYNAMICS

Assuming the $(C_{\text{SI}} \ln n)$ -spectral independence of $\mu = \mu_{G,q}$ for $G \sim \mathcal{G}(n, d/n)$ where C_{SI} depends on η, d and q , we show how to derive Theorem 1.1, the rapid mixing of Glauber dynamics P^{GD} for $\mu_{G,q}$. We follow the *entropic factorization* in [CLV26] with the degree separation on random graphs $\mathcal{G}(n, d/n)$ as in [BGŠ24] to obtain the rapid mixing of the block dynamics. To complete the mixing time of Glauber dynamics, we compare it with the block dynamics by the Dirichlet form.

5.1. Preliminaries. We first provide some necessary preliminaries.

Lemma 5.1 (Connectivity of list-colorings). *Let H be a finite graph with color lists $L(x)$. Suppose that its vertices admit an ordering $x_1, \dots, x_m$ such that, writing*

$$H_i := H[\{x_i, \dots, x_m\}],$$

we have

$$|L(x_i)| \geq \deg_{H_i}(x_i) + 2 \quad \text{for every } i.$$

Then proper list-colorings exist, and any two proper list-colorings can be joined by a sequence of proper list-colorings in which consecutive colorings differ at exactly one vertex.

Proof. Existence follows by coloring the vertices in reverse order.

We prove connectivity by induction on m . The assertion is immediate for $m \leq 1$. Let σ and τ be two proper list-colorings. By induction, their restrictions to $H - x_1$ can be joined by single-site recolorings.

We lift this sequence to H . Whenever a proposed recoloring of a neighbor of x_1 would assign the current color of x_1 to that neighbor, first recolor x_1 . Choose its new color from $L(x_1)$ so that it avoids all current neighbor colors and the proposed new color. At most $\deg_H(x_1) + 1$ colors are forbidden, whereas $|L(x_1)| \geq \deg_H(x_1) + 2$. Thus such a recoloring is always possible.

After lifting the entire sequence, all vertices other than x_1 agree with τ . Finally recolor x_1 to $\tau(x_1)$ if necessary. This last step is proper because τ is proper. $\square$

Markov chains and continuous generators. We briefly introduce the concept of Markov chains here and recommend to refer [LP17]. For a state space $\Omega \subseteq [q]^V$, a discrete-time Markov chain consists of a transition kernel P on Ω with its stationary distribution μ . Its transition process $X_0, \dots, X_t, \dots$ is generated by $X_{t+1} \sim P(X_t, \cdot)$ for every $t \in \mathbb{N}$. The mixing time of P is defined as

$$\tau_{\text{mix}}(P, \varepsilon) := \inf \{t \geq 0 : \forall x \in \Omega, \|P^t(x, \cdot) - \mu\|_{\text{TV}} \leq \varepsilon\}$$

In this paper, we only consider the *irreducible and self-adjoint* Markov chains.

A continuous Markov chain is specified by a generator $\mathcal{L}$ with non-negative off-diagonal entries and zero row sum. To test it, for every function $f : \Omega \rightarrow \mathbb{R}$, we observe the following function

$$(8) \quad (\mathcal{L}f)(x) := \sum_{y \in \Omega} \mathcal{L}(x, y)(f(y) - f(x)).$$

For simplicity, we sometimes use (8) to specify the generator $\mathcal{L}$. The transition semigroup is denoted by $H_t := e^{t\mathcal{L}}$, meaning that

$$(H_t f)(x) = \mathbb{E}[f(X_t) \mid X_0 = x], \quad \frac{d}{dt} H_t f|_{t=0} = \mathcal{L}f.$$

Given a discrete-time Markov chain P , its continuous-time version is the generator $\mathcal{L} = P - I$ obtained by applying a P -update at times of a rate-one Poisson clock. We always use the uppercase letter P to denote a discrete-time Markov chain and use the calligraphic letter $\mathcal{L}$ to denote the continuous-time generator.

Dirichlet form and Markov chain comparison. The *Dirichlet form* of a discrete-time Markov chain P is the following functional defined as: for all observable functions $f, g : \Omega \rightarrow \mathbb{R}$,

$$\mathcal{E}_P(f, g) := \frac{1}{2} \sum_{x, y \in \Omega} \mu(x) P(x, y) (f(x) - f(y))(g(x) - g(y)).$$

Meanwhile, for a continuous-time generator $\mathcal{L}$, the Dirichlet form is of the following form:

$$\mathcal{E}_{\mathcal{L}}(f, g) := \frac{1}{2} \sum_{x, y \in \Omega} \mu(x) \mathcal{L}(x, y) (f(x) - f(y))(g(x) - g(y)).$$

The following Rayleigh–Ritz variational characterization relates the Dirichlet form with the spectral gap:

$$\lambda_{\text{gap}}(P) := 1 - \lambda_2(P) = \inf_{\text{Var}_{\mu}(f) > 0} \frac{\mathcal{E}_P(f, f)}{\text{Var}_{\mu}(f)}, \quad \lambda_{\text{gap}}(\mathcal{L}) := -\lambda_2(\mathcal{L}) = \inf_{\text{Var}_{\mu}(f) > 0} \frac{\mathcal{E}_{\mathcal{L}}(f, f)}{\text{Var}_{\mu}(f)}.$$

Set the relaxation time $T_{\text{rel}}(\mathcal{M}) := \lambda_{\text{gap}}(\mathcal{M})^{-1}$ with convention $1/0 = \infty$.

The next comparison is between the single-site Glauber dynamics P^{GD} and the heat-bath block dynamics $\mathcal{L}_{\mathcal{B}}$. For a subset $S \subseteq V$, define its block heat-bath operator as

$$(\mathcal{K}_S f)(x) := \mathbb{E}_{\mu}[f(X) \mid X(V \setminus S) = x(V \setminus S)].$$

For a block cover $\mathcal{B} = \{B_1, \dots, B_r\}$ of V ($\cup_{B \in \mathcal{B}} B = V$) with update rates $w : \mathcal{B} \rightarrow \mathbb{R}_{>0}$, the continuous-time heat-bath block dynamics is specified by the generator $\mathcal{L}_{\mathcal{B}}$ defined as

$$\mathcal{L}_{\mathcal{B}} := \sum_{B \in \mathcal{B}} w(B) (\mathcal{K}_B - I).$$

The single-site Glauber dynamics P^{GD} is a special case of the heat-bath block dynamics: let $\mathcal{B}$ be the natural single-vertex cover of V and set $w(v) = 1/|V|$ for every vertex $v \in V$. Then the transition kernel of the discrete-time single-site Glauber dynamics is

$$P^{\text{GD}} = \frac{1}{|V|} \sum_{v \in V} \mathcal{K}_v.$$

The following comparison lemma builds a bridge between P^{GD} and the block dynamics $\mathcal{L}_{\mathcal{B}}$.

Lemma 5.2 ([Mar99, LM11]). *Given a block cover $\mathcal{B} = \{B_1, \dots, B_r\}$ with update rates $w : \mathcal{B} \rightarrow \mathbb{R}_{>0}$, for every $i = 1, \dots, r$ and every feasible partial assignment τ on $V \setminus B_i$, define the generator:*

$$(\mathcal{L}_i^{\tau} f)(x_{B_i}) := \sum_{v \in B_i} (\mathbb{E}_{\mu^{\tau}}[f(X_{B_i}) \mid X(B_i \setminus \{v\}) = x(B_i \setminus \{v\})] - f(x_{B_i})).$$

If there is a factor $R > 0$ such that for every $i = 1, \dots, r$ and every feasible partial assignment τ on $V \setminus B_i$, $T_{\text{rel}}(\mathcal{L}_i^\tau) \leq R$, then it holds that

$$T_{\text{rel}}(P^{\text{GD}}) \leq nR \cdot T_{\text{rel}}(\mathcal{L}_{\mathcal{B}}) \cdot \max_{v \in V} \sum_{B \in \mathcal{B}: v \in B} w_B.$$

The following lemma shows how to derive the mixing time from the relaxation time.

Lemma 5.3 ([LP17]). *For a discrete-time Markov chain P with its stationary distribution μ , it holds that*

$$\tau_{\text{mix}}(P, \varepsilon) \leq T_{\text{rel}}(P) \left(\ln \frac{1}{\min_{x: \mu(x) > 0} \mu(x)} + \ln \frac{1}{\varepsilon} \right).$$

Entropic factorization. Given a state space $\Omega \subseteq [q]^V$ and a probability distribution μ on Ω , for an observable function $f : \Omega \rightarrow \mathbb{R}$, the expectation of f with respect to μ is denoted by $\mathbb{E}_\mu[f] := \sum_{x \in \Omega} \mu(x) f(x)$ and its variance is $\mathbf{Var}_\mu(f) := \mathbb{E}_\mu[f^2] - \mathbb{E}_\mu[f]^2$. If f is non-negative valued, define its entropy as

$$\mathbf{Ent}_\mu[f] := \mathbb{E}_\mu[f \ln f] - \mathbb{E}_\mu[f] \ln \mathbb{E}_\mu[f]$$

with convention $0 \ln 0 = 0$. For $S \subseteq V$, define $\mathbf{Ent}_\mu^S(f) = \mathbb{E}_{\tau \sim \mu_{V \setminus S}}[\mathbf{Ent}_{\mu^\tau}[f]]$. The following inequality holds immediately for the tensor product.

Fact 5.4 ([BGGŠ24, Fact 15]). *Let $\mu = \mu_1 \otimes \dots \otimes \mu_k$ be the product distribution on $\Omega = \Omega_1 \times \dots \times \Omega_k$ where μ_i is a distribution on $[q]^{V_i}$ and $V_1, \dots, V_k$ are pairwise disjoint sets. Then for every test function $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$, $\mathbf{Ent}_\mu[f] \leq \sum_{i=1}^k \mathbf{Ent}_{\mu_i}^{V_i}(f)$.*

Remark 5.5. We remark here that the same inequality also holds for $\mathbf{Var}_\mu(f)$.

A distribution μ on $\Omega \subseteq [q]^V$ is said to be b -marginally bounded if for every boundary $\Lambda \subseteq V$, every feasible partial assignment τ on Λ , $v \in V \setminus \Lambda$ and every color $c \in [q]$, either $\mu^\tau(v \leftarrow c) = 0$ or $\mu^\tau(v \leftarrow c) \geq b$. A distribution is said to have *connected conditional supports* if, under every feasible pinning, its support is connected in the graph whose edges join configurations differing at exactly one free coordinate.

We use the following consequence of [CLV26, Lemma 4.2], with its connectivity assumption made explicit.

Lemma 5.6 ([CLV26, Lemma 4.2]). *Let μ be a probability distribution on $k \geq 1$ coordinates. Suppose that μ has connected conditional supports and is b -marginally bounded for some $b \in (0, 1/2]$. Then, for every non-negative function f ,*

$$\mathbf{Ent}_\mu[f] \leq 4b^{-(3k+1)} \sum_v \mathbf{Ent}_\mu^v(f).$$

Proof. For $k = 1$, the assertion follows from $\mathbf{Ent}_\mu[f] = \mathbf{Ent}_\mu^v(f)$.

For $k \geq 2$, the bound in [CLV26, Lemma 4.2] gives

$$\mathbf{Ent}_\mu[f] \leq \frac{3k^2 \ln(1/b)}{2b^{2k+2}} \sum_v \mathbf{Ent}_\mu^v(f).$$

The elementary inequality

$$\frac{3k^2 \ln(1/b)}{2b^{2k+2}} \leq 4b^{-(3k+1)}, \quad k \geq 2, \quad 0 < b \leq \frac{1}{2},$$

completes the proof. □

The following theorem shows that spectral independence together with marginal boundedness implies the entropic factorization.

Theorem 5.7 ([CLV26]). *For a probability distribution μ on $\Omega \subseteq [q]^V$, suppose that μ is γ -spectrally independent and b -marginally bounded with connected conditional supports. Then for every $r = 0, \dots, n$, there exists a factor $C_r > 0$ such that*

$$(9) \quad \frac{r}{|V|} \mathbf{Ent}_\mu[f] \leq C_r \frac{1}{\binom{|V|}{r}} \sum_{S \in \binom{V}{r}} \mathbf{Ent}_\mu^S(f).$$

Furthermore, if $r \geq 2R_0$ where $R_0 = \lceil 4\gamma/b^2 \rceil$, one can take

$$C_r \leq \left(\frac{|V| - R_0}{r - R_0} \right)^{R_0} \leq \left(\frac{2|V|}{r} \right)^{R_0}.$$

Remark 5.8. In [CLV26], the precise expression of C_r is

$$C_r = \frac{r}{|V|} \frac{\sum_{k=0}^{|V|-1} \Gamma_k}{\sum_{k=|V|-r}^{|V|-1} \Gamma_k}, \quad \Gamma_k = \prod_{j=0}^{k-1} \max \left\{ 0, 1 - \frac{4\gamma}{b^2(|V| - 1 - j)} \right\}.$$

For convenience, we use the explicit form in Theorem 5.7.

The following proposition provides that the entropic factorization implies an upper bound for the relaxation time.

Proposition 5.9 ([CLV26, Fact 3.5]). *If $\mathbf{Ent}_\mu[f] \leq C \sum_{v \in V} \mathbf{Ent}_\mu^v(f)$ for every observable f , then the following Poincaré inequality holds:*

$$\mathbf{Var}_\mu(g) \leq C \sum_{v \in V} \mathbb{E}_\mu[\mathbf{Var}_\mu(g \mid X(V \setminus \{v\}))] = C \mathcal{E}_\mathcal{L}(g, g)$$

where $\mathcal{L}$ is the corresponding continuous-time single-site Glauber dynamics of μ , i.e.

$$\mathcal{L} = \sum_{v \in V} (\mathcal{K}_v - I).$$

Thus $T_{\text{rel}}(\mathcal{L}) \leq C$.

5.2. High-degree buffers. Now we show how to handle the high-degree vertices in $G \sim \mathcal{G}(n, d/n)$. Recall the high-degree cutoff $D = \lfloor (1 + \eta/3)d \rfloor$ and the high-degree vertex set

$$V^{\text{HD}} = \{v \in V \mid \deg_G(v) > D\}.$$

Define $S := N_G[V^{\text{HD}}]$ and $U := V \setminus S$. Take d sufficiently large such that $q - D \geq 3$, $q \geq 5$ and $D - 6 > d$.

The following buffered percolation accommodates the dependence between high-degree events.

Lemma 5.10 (Buffered percolation). *There is a fixed real number $p \in (0, 1)$ depending on η and d such that the following holds. Let Z mark each vertex independently with probability p independently of G and set $A = V^{\text{HD}} \cup Z$. For every fixed $J > 0$, there is a $K_J < \infty$ such that*

$$\mathbb{P}_{G, Z} \left(\max_{R \text{ component of } G[N_G[A]]} |R| > K_J \ln n \right) \leq n^{-J}$$

for all sufficiently large n . Moreover, with probability at least $1 - O_{\eta, d}(n^{-1})$ over G every component of $G[S]$ is a tree or unicyclic and has at most $K_1 \ln n$ vertices.

Proof. For a fixed edge set F , by Lemma 3.4,

$$\mathbb{P}(B \subseteq V^{\text{HD}} \mid F \subseteq G) \leq e^{2t|F|} a_0^{|B|} \quad \text{where} \quad a_0 = \exp\left(-tD + \frac{d}{2}(e^{2t} - 1)\right).$$

Therefore,

$$\begin{aligned} \mathbb{P}(B \subseteq A \mid F \subseteq G) &\leq \sum_{C \subseteq B} p^{|B|-|C|} \mathbb{P}(C \subseteq V^{\text{HD}} \mid F \subseteq G) \\ &\leq e^{2t|F|} (a_0 + p)^{|B|}. \end{aligned}$$

Define

$$\begin{aligned} \psi(x) &:= x \ln x - x + 1, \quad x = \frac{D-6}{d}, \quad t = \frac{1}{2} \ln x, \\ \rho &:= e^{6t} (a_0 + p) = e^{-\frac{d}{2}\psi((D-6)/d)} + p e^{6t}. \end{aligned}$$

For sufficiently large d , pick p small enough such that

$$\theta_0 := (2ed)^3 \rho < \frac{1}{16}.$$

The choice of p is possible and only depends on η and d since p decays exponentially in d for fixed η .

Conditional on a revealed pair of (G, Z) , we say the vertices in A are *seeds*. For every non-seed vertex in $N_G[A]$, we assign it to one adjacent seed using a fixed tie-breaking rule. Additionally each seed owns itself. The assigned edges together with some seeds themselves form disjoint stars covering $N_G[A]$. We contract all star edges and keep other edges including loops and parallel edges. If a component contains at least j seeds, a tree on exactly j vertices can be selected in the quotient graph after contraction. Now we lift this quotient tree back to G , including its j center seeds, the original edges representing its $j-1$ quotient edges and the star arms needed to connect their endpoints to their centers.

We count the number of original vertices after lifting. Observe that a quotient edge introduces at most two non-seed vertices. Then the lifted certificate is connected on $j \leq k \leq 3j-2$ vertices. Note that there are exactly $k-j$ star arms in this lifting graph with $j-1$ inter-star edges. Then it has $k-1$ edges, leading to the fact that it is a tree together with connectivity. For such a certificate B , by Lemma 3.4,

$$\mathbb{P}(B \subseteq A \mid F \subseteq G) \leq e^{2t(k-1)} (a_0 + p)^j \leq \rho^j$$

since $k-1 \leq 3j$. Then with Cayley's formula, the expected number of certificates is at most

$$\sum_{k=j}^{3j} \binom{n}{k} k^{k-2} 2^k \left(\frac{d}{n}\right)^{k-1} \rho^j \leq O_d(n(j+1)\theta_0^j).$$

Fix $J > 0$ and we pick L_J so that $L_J \ln(1/\theta) > J + 4$. Markov inequality provides that for all sufficiently large n ,

$$\mathbb{P}(\text{some component contains at least } L_J \ln n \text{ seeds}) \leq n^{-J-2}.$$

Now we bound the neighborhood of seeds. Fix a vertex set W of size k and a tree F on W . Conditional on $F \subseteq G$, a similar calculation as Lemma 3.4 shows, for any fixed $s > 0$,

$$\mathbb{E} \left[e^{s \sum_{v \in W} \deg_G(v)} \mid F \subseteq G \right] \leq \exp \left(2s(k-1) + \frac{kd}{2} (e^{2s} - 1) \right).$$

Consequently by Markov inequality,

$$\mathbb{P}\left(F \subseteq G, \sum_{v \in W} \deg_G(v) > M \ln n\right) \leq \left(\frac{d}{n}\right)^{k-1} n^{-sM} \exp\left(2s(k-1) + \frac{kd}{2}(e^{2s} - 1)\right).$$

Set

$$B_d := ed \exp\left(2s + \frac{d}{2}(e^{2s} - 1)\right) > 1.$$

Union over all spanning trees and $k = 1, \dots, 3L_J \ln n$, and we obtain that

$$\mathbb{P}\left(\exists W \subseteq V, G[W] \text{ connected}, |W| \leq 3L_J \ln n, \sum_{v \in W} \deg_G(v) > M \ln n\right) \leq n^{1-sM+3L_J \ln B_d + o(1)}.$$

Pick an arbitrary fixed $s > 0$ and choose M large enough such that $sM > 3L_J \ln B_d + J + 4$. Then the probability is at most n^{-J-2} .

Now, suppose that every component of $G[N_G[A]]$ contains at most $K_J \ln n$ seeds, and for every connected vertex subset W of size at most $3L_J \ln n$, $\sum_{v \in W} \deg_G(v) \leq M \ln n$. For every connected component R in $G[N_G[A]]$, let $B_R = A \cap R$ and $r = |B_R| \leq L_J \ln n$. Lift its quotient spanning tree connecting all its seeds. Its vertex set W is connected and contains at most $3r \leq 3L_J \ln n$ vertices. Therefore,

$$|R| \leq |B_R| + \sum_{v \in W} \deg_G(v) \leq (L_J + M) \ln n.$$

Take $K_J := L_J + M$ which depends only on η, d and J and we conclude the first argument.

To show the second argument, for a connected graph Q , define its cycle rank:

$$\text{CycleRank}(Q) := |E(Q)| - |V(Q)| + 1.$$

We use the star partition and the contraction again with seeds V^{HD} only and thus its vertex set is exactly $S = N_G[V^{\text{HD}}]$. Note that the contraction does not decrease the cycle rank if we keep loops and parallel edges. Then if Q has at least two cycles ($\text{CycleRank}(Q) \geq 2$), its quotient also does. Suppose that the quotient contains j vertices. We pick a spanning tree of it and two additional edges to obtain a certificate on j vertices with $j + 1$ edges.

Suppose that a component Q of $G[S]$ has $\text{CycleRank}(Q) \geq 2$. Contract the stars centered at the seeds V^{HD} , retaining loops and parallel edges. Contraction does not decrease the cycle rank. Hence the quotient contains a spanning tree together with two additional edges.

Suppose that this quotient certificate has j seed vertices. Lifting it back to G gives a connected graph F on k vertices with

$$j \leq k \leq 3j, \quad |E(F)| = k + 1,$$

and a set $B \subseteq V(F)$ of j vertices satisfying $B \subseteq V^{\text{HD}}$. By Lemma 3.4,

$$\mathbb{P}(B \subseteq V^{\text{HD}} \mid F \subseteq G) \leq e^{2t(k+1)} a_0^j.$$

Therefore the expected number of such certificates is at most

$$\begin{aligned} & \sum_{j \geq 1} \sum_{k=j}^{3j} \binom{n}{k} k^{k-2} k^4 2^k \left(\frac{d}{n}\right)^{k+1} e^{2t(k+1)} a_0^j \\ & \leq \frac{C_d}{n} \sum_{k \geq 1} k^3 ((2ed)^3 e^{6t} a_0)^{k/3} = O_{\eta, d}(n^{-1}), \end{aligned}$$

because

$$(2ed)^3 e^{6t} a_0 \leq (2ed)^3 \rho < \frac{1}{16}.$$

Thus, with probability $1 - O_{\eta,d}(n^{-1})$, every component of $G[S]$ is a tree or unicyclic. Combining this with the first part of the proof gives the asserted $K_1 \ln n$ component-size bound. Put everything together and we conclude the lemma. $\square$

5.3. Buffered marginals. Now we study the marginal distribution on U . Let $\nu = \mu_U$ be the marginal distribution of $\mu = \mu_{G,q}$ projected on U . Assume the event that every component in $G[S]$ is a tree or unicyclic of size at most $K_1 \ln n$.

We first state the following marginal lower bound for ν .

Proposition 5.11. *With high probability, $|U| \geq c_d n$ where $c_d = e^{-d}/2$. Every proper coloring of $G[U]$ extends to a proper coloring of G . Furthermore, ν is b -marginally bounded under all its feasible pinnings with*

$$b = q^{-(D+1)}.$$

Proof. The extension argument is a direct corollary of [Yin14, Lemma 3.4]. Let

$$\partial U := \{x \in V \setminus U : N_G(x) \cap U \neq \emptyset\}$$

be the external vertex boundary of U . We first verify that U is a permissive block in the full-list instance $\mathcal{L}(x) = [q]$. Indeed, $\partial U \cap V^{\text{HD}} = \emptyset$. Otherwise, if $x \in \partial U \cap V^{\text{HD}}$, then

$$N_G(x) \subseteq N_G[V^{\text{HD}}] = S,$$

contradicting $N_G(x) \cap U \neq \emptyset$, since $U = V \setminus S$. Consequently, every $x \in \partial U$ satisfies

$$\deg_G(x) \leq D \quad \text{and} \quad |\mathcal{L}(x)| = q \geq D + 3 > \deg_G(x) + 1.$$

Thus U is a permissive block. Applying [Yin14, Lemma 3.4] with $B = U$ and $\Delta = \emptyset$, every proper coloring of $G[U]$ is globally feasible, and hence extends to a proper coloring of G .

Now we show the marginal boundedness of ν . Fix a feasible pinning τ on $P \subseteq U$, a free vertex $v \in U \setminus P$ and a color c which is feasible at v . Define

$$\Omega^\tau := \{\sigma \in \Omega(G, q) : \sigma(P) = \tau\}, \quad \Omega^{\tau,c} := \{\sigma \in \Omega^\tau : \sigma(v) = c\}.$$

Then

$$\nu^\tau(v \leftarrow c) = \frac{|\Omega^{\tau,c}|}{|\Omega^\tau|}.$$

We construct a mapping from Ω^τ to $\Omega^{\tau,c}$ as follows. Start from $\sigma \in \Omega^\tau$. check the neighbors of v in a prescribed order. Whenever a neighbor u currently is assigned the color c , recolor it to the smallest color which is not c and also not used by its neighbors. After the exploration, change the color of v to c . Therefore, for any output $\pi \in \Omega^{\tau,c}$, there are at most $q^{|N_G[v]|}$ pre-images in Ω^τ corresponding to π , leading to the bound

$$|\Omega^\tau| \leq q^{|N_G[v]|} |\Omega^{\tau,c}| \leq q^{D+1} |\Omega^{\tau,c}|.$$

This bound concludes the marginal boundedness of ν .

To see $|U| \geq e^{-d}n/2$, it is convenient to calculate the number of isolated vertices in G . A typical calculation shows that with probability $1 - o_n(1)$, G contains at least $e^{-d}n/2$ isolated vertices, and thus $|U| \geq c_d n$ for $c_d = e^{-d}/2$. For simplicity we omit the calculation here and suggest to refer [FK15]. $\square$

Corollary 5.12 (Connectivity of buffered marginals). *Assume the extension property in Proposition 5.11 and $q - D \geq 2$. Then $\nu = \mu_U$ has connected conditional supports. Moreover, every coordinate marginal of a feasible conditional distribution of ν also has connected conditional supports.*

Proof. Fix a feasible pinning τ on $P \subseteq U$, and put $F = U \setminus P$. For each $x \in F$, define

$$L_\tau(x) := [q] \setminus \{\tau(y) : y \in N_G(x) \cap P\}.$$

By the extension property, the support of ν^τ is exactly the set of proper list-colorings of $G[F]$ with lists L_τ . Although ν^τ need not be uniform, this identification of its support is sufficient.

Since every vertex of U has degree at most D ,

$$\begin{aligned} |L_\tau(x)| &\geq q - |N_G(x) \cap P| \\ &= \deg_{G[F]}(x) + q - \deg_{G[U]}(x) \\ &\geq \deg_{G[F]}(x) + q - D \\ &\geq \deg_{G[F]}(x) + 2. \end{aligned}$$

Thus any vertex ordering satisfies the hypothesis of Lemma 5.1. Hence the support of ν^τ is connected. Since P and τ were arbitrary, ν has connected conditional supports.

Finally, the coordinate projection of a connected single-site support is connected: projecting a single-site path gives a path after consecutive repetitions are removed. Applying this observation after any additional feasible pinning proves the assertion about marginals. $\square$

The second ingredient on the buffered probability distribution ν is the spectral independence of it.

Proposition 5.13. *Let μ be a distribution on $[q]^V$ and $\nu = \mu_U$ for $U \subseteq V$. If μ is γ -spectrally independent under every feasible pinning, then ν is also γ -spectrally independent.*

Proof. Observe that for every feasible pinning τ on $\Lambda \subseteq U$, Ψ_ν^τ is a principal submatrix of Ψ_μ^τ . After symmetrizing Ψ_μ^τ and by Cauchy interlacing theorem, it is easy to show

$$\lambda_{\max}(\Psi_\nu^\tau) \leq \lambda_{\max}(\Psi_\mu^\tau) \leq \gamma.$$

Thus we conclude this proposition. $\square$

5.4. Entropic factorization. Now we establish the entropic factorization for ν . Pick p as in Lemma 5.10, $\zeta = p/4$ and set $r = \lfloor \zeta|U| \rfloor$, $\vartheta = r/|U|$. For sufficiently large n , $\zeta/2 \leq \vartheta \leq \zeta$ and $r \geq 2R_0$. Pick

$$a := \left(\frac{4C_{\text{SI}}}{b^2} + 2 \right) \ln(4/\zeta) + 2.$$

Lemma 5.14. *Fix $J > 2a + 6$ and take $K = K_J$ in Lemma 5.10. With high probability over $G \sim \mathcal{G}(n, d/n)$,*

$$\mathbb{P}_{T \sim \binom{U}{r}} \left(\max_{R \text{ component of } G[S \cup T]} |R| > K \ln n \mid G \right) \leq n^{-a-2}.$$

Proof. Conditional on G , couple the uniform subset $T \in \binom{U}{r}$ to Bernoulli- p marks on U as follows: draw a permutation $u_1, \dots, u_{|U|}$ of U uniformly at random and independently draw $B \sim \text{Binomial}(|U|, p)$. Set

$$T = \{u_1, \dots, u_r\}, \quad Z_U = \{u_1, \dots, u_B\}.$$

It is clear that T is a uniform r -subset of U and Z_U follows the Bernoulli- p marks on U . Additionally, mark the vertices of S independently with probability p . Let Z be the full marked set. Then for every $W \subseteq V$,

$$\mathbb{P}(Z = W \mid G) = p^{|W|} (1-p)^{n-|W|}.$$

Since $\vartheta = r/|U| \leq \zeta = p/4$ and $\mathbb{E}[B \mid G] = p|U|$, the Chernoff bound provides that

$$\mathbb{P}(B < r \mid G) \leq \exp\left(-\frac{9p|U|}{32}\right).$$

On the event $B \geq r$, it holds that $T \subseteq Z_U$ and hence $T \subseteq Z$, meaning that $S \cup T \subseteq N_G[V^{\text{HD}} \cup Z]$. Therefore, conditional on the event $B \geq r$, the event a component of $G[S \cup T]$ surpassing $K \ln n$ forces the event that a component of $G[N_G[V^{\text{HD}} \cup Z]]$ is larger than $K \ln n$.

Denote by $\mathcal{E}$ the event that there is a component of $G[S \cup T]$ of size larger than $K \ln n$. By Lemma 5.10, it holds that

$$\begin{aligned} \mathbb{E}[\mathbb{P}(\mathcal{E}) \cdot \mathbb{1}[|U| \geq c_d n]] &\leq n^{-J} + \mathbb{E}_G[\mathbb{1}[|U| \geq c_d n] \cdot \mathbb{P}(B < r \mid G)] \\ &\leq n^{-J} + \exp\left(-\frac{9pc_d n}{32}\right). \end{aligned}$$

By Markov inequality,

$$\mathbb{P}_G(\mathbb{P}(\mathcal{E}) > n^{-a-2} \wedge \mathbb{1}[|U| \geq c_d n]) \leq n^{a+2-J} + n^{a+2} \exp\left(-\frac{9pc_d n}{32}\right)$$

Union bound together with Proposition 5.11 provides that

$$\mathbb{P}_G(\mathbb{P}(\mathcal{E}) > n^{-a-2}) \leq o_n(1) + n^{a+2-J} + n^{a+2} \exp\left(-\frac{9pc_d n}{32}\right)$$

which is still $o_n(1)$ for sufficiently large n . Thus we conclude the lemma. $\square$

Now we are ready to establish the following entropic factorization for ν .

Proposition 5.15. *With high probability over $G \sim \mathcal{G}(n, d/n)$, for every observable test function f ,*

$$\mathbf{Ent}_\nu[f] \leq C_{\text{LD}} \sum_{u \in U} \mathbf{Ent}_\nu^u(f), \quad C_{\text{LD}} = 2C_r \cdot 4b^{-(3K \ln n + 1)}$$

where K is decided by Lemma 5.14. Consequently, the continuous-time marginal Glauber dynamics $\mathcal{L}^\nu$ with rate one per vertex of U has relaxation time at most $C_{\text{LD}} \leq n^B$ for a fixed constant $B > 0$ depending only on η, d and q .

Proof. Assume the joint event of Proposition 5.11 and Lemma 5.14. In addition, assume that $\mu = \mu_{G,q}$ is $(C_{\text{SI}} \ln n)$ -spectrally independent for some constant $C_{\text{SI}} = C_{\text{SI}}(\eta, d, q)$ and therefore ν is $(C_{\text{SI}} \ln n)$ -spectrally independent. By Corollary 5.12, ν has connected conditional supports. Together with its marginal boundedness and spectral independence, Theorem 5.7 gives

$$(10) \quad \vartheta \cdot \mathbf{Ent}_\nu[f] \leq C_r \mathbb{E}_{T \sim \binom{U}{r}}[\mathbf{Ent}_\nu^T(f)]$$

where $C_r \leq n^a$ (recall that $a = (4C_{\text{SI}}/b^2 + 2) \ln(4/\zeta) + 2$). For simplicity, we say an r -subset T of U is good if every component of $G[S \cup T]$ has at most $K \ln n$ vertices and otherwise it is bad. By Lemma 5.14, $\mathbb{P}_T(T \text{ is bad}) \leq n^{-a-2}$.

Now we fix a good subset T and a feasible assignment τ on $U \setminus T$. Let R range over all connected components of $G[S \cup T]$ and set $A_R = R \cap T$. Then it holds that

$$\nu^\tau = \bigotimes_{R: A_R \neq \emptyset} \nu(A_R \leftarrow \cdot \mid \tau).$$

For every R , since T is good, by Proposition 5.11, the following properties hold for $\nu(A_R \leftarrow \cdot \mid \tau)$:

- it has at most $K \ln n$ coordinates since $|A_R| \leq |R| \leq K \ln n$;
- it has connected conditional supports by Corollary 5.12;
- it is b -marginally bounded for $b = q^{-(D+1)}$.

By Lemma 5.6,

$$\mathbf{Ent}_{\nu(A_R \leftarrow \cdot \mid \tau)}[f] \leq 4b^{-(3K \ln n + 1)} \sum_{u \in A_R} \mathbf{Ent}_{\nu(A_R \leftarrow \cdot \mid \tau)}^u(f).$$

Then by Fact 5.4,

$$\mathbf{Ent}_{\nu^\tau}[f] \leq 4b^{-(3K \ln n + 1)} \sum_{u \in T} \mathbf{Ent}_{\nu^\tau}^u(f).$$

Taking expectation over both sides, by the tower property of expectation, for every good r -subset T ,

$$\mathbf{Ent}_\nu^T(f) \leq 4b^{-(3K \ln n + 1)} \sum_{u \in T} \mathbf{Ent}_\nu^u(f).$$

For a bad r -subset T , the chain rule of entropy shows

$$\mathbf{Ent}_\nu[f] = \mathbf{Ent}_\nu^T(f) + \mathbf{Ent}_\nu[\mathbb{E}_\nu[f \mid X(U \setminus T)]]$$

which implies that $\mathbf{Ent}_\nu^T(f) \leq \mathbf{Ent}_\nu[f]$ by non-negativity.

Back to (10), by Lemma 5.14,

$$\vartheta \cdot \mathbf{Ent}_\nu[f] \leq C_r \left(\vartheta \cdot 4b^{-(3K \ln n + 1)} \sum_{u \in U} \mathbf{Ent}_\nu^u(f) + n^{-a-2} \cdot \mathbf{Ent}_\nu[f] \right).$$

For sufficiently large n , $C_r n^{-a-2} \leq n^{-2} \leq \vartheta/2$. Therefore,

$$\mathbf{Ent}_\nu[f] \leq 2C_r \cdot 4b^{-(3K \ln n + 1)} \sum_{u \in U} \mathbf{Ent}_\nu^u(f)$$

which is the desired entropic factorization. With the entropic factorization for ν and the choice of $B = a + 3K \ln(1/b) + 3$, the relaxation time is a direct corollary of Proposition 5.9. $\square$

5.5. Comparison back to the original chain. Now we put back our attention to the original distribution μ and its corresponding discrete-time Glauber dynamics P^{GD} . For every $u \in U$, define $B_u = S \cup \{u\}$. Then the block cover $\mathcal{B}$ is defined as $\mathcal{B} = \{B_u : u \in U\}$. Define the operator P_u as: for every test function f , $P_u f = \mathbb{E}_\mu[f \mid X(U \setminus \{u\})]$. Hence the generator and its Dirichlet form are set by

$$\mathcal{L}_\mathcal{B} := \sum_{u \in U} (P_u - I), \quad \mathcal{E}_{\mathcal{L}_\mathcal{B}}(f, f) = \sum_{u \in U} \mathbb{E}_\mu[\mathbf{Var}(f \mid X(U \setminus \{u\}))].$$

Lemma 5.16. *The following identity holds for this block dynamics:*

$$\lambda_{\text{gap}}(\mathcal{L}_\mathcal{B}) = \min \{|U|, \lambda_{\text{gap}}(\mathcal{L}^\nu)\}$$

where $\mathcal{L}^\nu$ is the continuous-time single-site Glauber dynamics for ν , with the $|U|$ branch omitted if its invariant subspace is zero. Consequently, its relaxation time is at most $T_{\text{rel}}(\mathcal{L}_\mathcal{B}) \leq C_{\text{LD}}$ assuming Proposition 5.15.

Proof. We decompose the space $L^2(\mu)$. Define the operator Q as for every $f \in L^2(\mu)$,

$$Qf = \mathbb{E}_\mu[f \mid X(U)].$$

Consider the two subspace

$$\mathcal{F} := \{f \in L^2(\mu) : f(x) = f(y) \text{ whenever } x(U) = y(U)\}, \quad \mathcal{H} := \{h : \mathbb{E}_\mu[h \mid X(U)] = 0\}$$

Therefore, for every $f \in L^2(\mu)$, f has a unique decomposition $f = Qf + (f - Qf)$ such that $Qf \in \mathcal{F}$ and $f - Qf \in \mathcal{H}$. It is easy to verify that for every $g \in \mathcal{F}$ and $h \in \mathcal{H}$, $\mathbb{E}_\mu[gh] = 0$. Then we represent $L^2(\mu) = \mathcal{F} \oplus \mathcal{H}$.

Since ν is the marginal of μ projected on U , the first identity is $\mathbb{E}_\mu [g(X(U))^2] = \mathbb{E}_\nu [g^2]$. Moreover, let $f(x) = g(x(U))$. It holds that

$$P_u f = \mathbb{E}_\mu [g(X(U)) \mid X(U \setminus \{u\})] = \mathbb{E}_\nu [g \mid X(U \setminus \{u\})].$$

Summing over u , we obtain that

$$\mathcal{L}_\mathcal{B}(g(X(U))) = (\mathcal{L}^\nu g)(X(U)).$$

For $h \in \mathcal{H}$, the tower property provides that

$$P_u h = \mathbb{E}_\mu [h \mid X(U \setminus \{u\})] = \mathbb{E}_\mu [\mathbb{E}_\mu [h \mid X(U)] \mid X(U \setminus \{u\})] = 0.$$

Hence $\mathcal{L}_\mathcal{B} h = -|U|h$. Then for every non-constant $f \in L^2(\mu)$ with decomposition $f = g + h$,

$$\frac{\mathcal{E}_{\mathcal{L}_\mathcal{B}}(f, f)}{\mathbf{Var}_\mu(f)} = \frac{\mathcal{E}_{\mathcal{L}^\nu}(g, g) + |U| \cdot \mathbb{E}_\mu [h^2]}{\mathbf{Var}_\nu(g) + \mathbb{E}_\mu [h^2]} \geq \min \{ \lambda_{\text{gap}}(\mathcal{L}^\nu), |U| \}.$$

To take the equality, only need to consider $\mathcal{F}$ and $\mathcal{H}$ respectively. Moreover, when $\mathcal{H} = \{0\}$, the $|U|$ is obviously omitted by calculation. Thus we conclude the spectral gap for $\mathcal{L}_\mathcal{B}$.

To show the relaxation time, it is clear that $T_{\text{rel}}(\mathcal{L}^\nu) \leq C_{\text{LD}}$. When $U \neq \emptyset$, by Propositions 5.9 and 5.15,

$$\mathbf{Var}_\nu(g) \leq C_{\text{LD}} \sum_{u \in U} \mathbb{E}_\nu [\mathbf{Var}(g \mid X(U \setminus \{u\}))] \leq C_{\text{LD}} |U| \cdot \mathbf{Var}_\nu(g).$$

Therefore $|U|^{-1} \leq C_{\text{LD}}$. Hence by Proposition 5.9,

$$T_{\text{rel}}(\mathcal{L}_\mathcal{B}) \leq \max \{ T_{\text{rel}}(\mathcal{L}^\nu), |U|^{-1} \} \leq C_{\text{LD}}.$$

Thus we conclude the full lemma. $\square$

The following proposition provides the relaxation time of Glauber dynamics restricted on B_u .

Proposition 5.17. *Assume the event in Lemma 5.10. For every block B_u and every feasible outside pinning τ on $V \setminus B_u$, the continuous-time single-site Glauber dynamics is connected and has relaxation time at most $q^{2(1+DK_1)\ln n+1}$ for all sufficiently large n .*

Proof. Fix a vertex $u \in U$ and a feasible pinning τ on $V \setminus B_u = U \setminus \{u\}$. We firstly consider the connected component containing u in $G[S \cup \{u\}]$. Note that u connects at most D components in $G[S]$. Therefore this component is of size at most $1 + DK_1 \ln n \leq (1 + DK_1) \ln n$ for sufficiently large n . Other components in $G[S]$ keep unchanged. Hence every component in $G[B_u]$ has at most $(1 + DK_1) \ln n$ vertices. We next prove connectivity under the outside pinning. Let $H = G[B_u]$, and for each $x \in B_u$ define

$$L_\tau(x) := [q] \setminus \{ \tau(y) : y \in N_G(x) \setminus B_u \}.$$

The conditional support is exactly the set of proper list-colorings of H with these lists.

First, H is 3-degenerate. Indeed, every component of $G[S]$ is a tree or unicyclic, so every nonempty induced subgraph of $G[S]$ has a vertex of degree at most 2. For any nonempty $A \subseteq B_u$ with $A \setminus \{u\} \neq \emptyset$, choose such a vertex in $G[A \setminus \{u\}]$. It has degree at most 3 in $G[A]$. The remaining case $A = \{u\}$ is immediate.

Choose a degeneracy ordering $x_1, \dots, x_m$ of H , and write

$$H_i = H[\{x_i, \dots, x_m\}], \quad \deg_{H_i}(x_i) \leq 3.$$

We verify the list-size condition in Lemma 5.1.

If $x_i \in V^{\text{HD}}$, then all neighbors of x_i belong to $S = N_G[V^{\text{HD}}] \subseteq B_u$. Thus no outside pinning removes a color from its list, and

$$|L_\tau(x_i)| = q \geq 5 \geq \deg_{H_i}(x_i) + 2.$$

If $x_i \notin V^{\text{HD}}$, then $\deg_G(x_i) \leq D$, and

$$\begin{aligned} |L_\tau(x_i)| &\geq q - |N_G(x_i) \setminus B_u| \\ &= \deg_H(x_i) + q - \deg_G(x_i) \\ &\geq \deg_H(x_i) + q - D \\ &\geq \deg_{H_i}(x_i) + 2. \end{aligned}$$

Therefore Lemma 5.1 applies. Every proper single-site recoloring has positive heat-bath transition rate, so the conditional continuous-time Glauber dynamics on B_u is irreducible.

To bound the relaxation time, for a component R of $G[S \cup \{u\}]$, let s be its size and Ω_R be its proper list-colorings. Write $N = |\Omega_R| \leq q^s$ and π be the conditional probability distribution of μ^τ on R . Observe that π is uniform over Ω_R . Let $\mathcal{L}_R$ be the generator of the continuous-time Glauber dynamics on Ω_R . Define the directed and undirected transition-edge sets by

$$\begin{aligned} \vec{\mathcal{T}}_R &:= \{(\sigma, \xi) \in \Omega_R^2 : \sigma \neq \xi, \mathcal{L}_R(\sigma, \xi) > 0\}, \\ \mathcal{T}_R &:= \{\{\sigma, \xi\} : (\sigma, \xi) \in \vec{\mathcal{T}}_R\}. \end{aligned}$$

For each $e = \{\sigma, \xi\} \in \mathcal{T}_R$, fix an arbitrary orientation and define

$$\Delta_e g := g(\sigma) - g(\xi).$$

For every ordered pair $(\sigma, \xi) \in \Omega_R^2$, choose a simple path $P_{\sigma, \xi}$ in the undirected transition graph. Since this path has length at most $N - 1$, the Cauchy–Schwarz inequality gives

$$\begin{aligned} \text{Var}_\pi(g) &= \frac{1}{2N^2} \sum_{\sigma, \xi \in \Omega_R} (g(\sigma) - g(\xi))^2 \\ &\leq \frac{1}{2N^2} \sum_{\sigma, \xi \in \Omega_R} (N - 1) \sum_{e \in P_{\sigma, \xi}} (\Delta_e g)^2 \\ &\leq \frac{N - 1}{2} \sum_{e \in \mathcal{T}_R} (\Delta_e g)^2. \end{aligned}$$

Since every off-diagonal transition rate is at least $1/q$,

$$\begin{aligned} \mathcal{E}_R(g, g) &= \frac{1}{2} \sum_{(\sigma, \xi) \in \vec{\mathcal{T}}_R} \frac{1}{N} \mathcal{L}_R(\sigma, \xi) (g(\sigma) - g(\xi))^2 \\ &\geq \frac{1}{2qN} \sum_{(\sigma, \xi) \in \vec{\mathcal{T}}_R} (g(\sigma) - g(\xi))^2 \\ &= \frac{1}{qN} \sum_{e \in \mathcal{T}_R} (\Delta_e g)^2. \end{aligned}$$

Therefore the following Poincaré inequality holds:

$$\text{Var}_\pi(g) \leq \frac{qN(N - 1)}{2} \mathcal{E}_R(g, g).$$

Thus the relaxation time for a single component R is at most

$$qN^2 \leq q^{1+2s} \leq q^{1+2(1+DK_1)\ln n}.$$

By Fact 5.4, we conclude the relaxation time for the continuous-time Glauber dynamics for $\mu_{B_u}^\tau$. $\square$

Proof of Theorem 1.1. Assume the joint event of Corollary 1.5, Lemmas 5.10 and 5.14, and Proposition 5.11. This joint event happens with probability $1 - o_n(1)$. Write P^{GD} as the discrete-time Glauber dynamics for μ , $\mathcal{L}^\mu$ as the continuous-time single-site chain with rate one clock per vertex of V , $\mathcal{L}_B$ as the above heat-bath block dynamics. By Lemma 5.2 with Propositions 5.15 and 5.17 and Lemma 5.16,

$$\text{Var}_\mu(f) \leq C_{\text{LD}} \mathcal{E}_{\mathcal{L}_B}(f, f) \leq C_{\text{LD}} \cdot n \cdot q^{1+2(1+DK_1)\ln n} \mathcal{E}_{\mathcal{L}^\mu}(f, f) \leq n^{C_1} \cdot \mathcal{E}_{P^{\text{GD}}}(f, f)$$

for the constant $C_1 = B + 2(1 + DK_1)\ln q + 10$ which depends only on η, d and q . Therefore $T_{\text{rel}}(P^{\text{GD}}) \leq n^{C_1}$. Note that for every feasible coloring $\sigma \in \Omega_{G,q}$, $\mu(\sigma) \geq q^{-n}$. Therefore by Lemma 5.3, P^{GD} exhibits mixing time at most $n^{C_1}(n \ln q + \ln(1/\varepsilon)) \leq n^C(1 + \ln(1/\varepsilon))$ for some constant C depending only on η, d and q . $\square$

REFERENCES

- [AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, pages 1198–1211, New York, NY, USA, 2020. Association for Computing Machinery. 3
- [ALOG24] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing*, 53(6):FOCS20–1–FOCS20–37, 2024. 3
- [AN05] Dimitris Achlioptas and Assaf Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162(3):1335–1351, 2005. 5
- [BCC⁺22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692, 2022. 3
- [BGGŠ24] Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, and Daniel Štefankovič. Fast sampling via spectral independence beyond bounded-degree graphs. *ACM Trans. Algorithms*, 20(1), January 2024. 3, 4, 24, 26
- [BS20] Louise Budzynski and Guilhem Semerjian. The asymptotics of the clustering transition for random constraint satisfaction problems. *Journal of Statistical Physics*, 181(5):1490–1522, 2020. 2
- [CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for Markov chains (extended abstract). In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122, 2022. 3
- [CF25] Xiaoyu Chen and Weiming Feng. Rapid mixing via coupling independence for spin systems with unbounded degree. In Alina Ene and Eshan Chattopadhyay, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025)*, volume 353 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 68:1–68:17, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 4
- [CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 588–599, 2022. 3
- [CGŠV21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557, 2021. 3

- [CLMM26] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications. *J. ACM*, August 2026. Just Accepted. 4, 5, 9, 10, 13, 14, 38
- [CLV23] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. *SIAM Journal on Computing*, 52(1):196–237, 2023. 3
- [CLV26] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. *SIAM Journal on Computing*, 55(4):STOC21–104–STOC21–153, 2026. 3, 4, 24, 26, 27
- [CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4478–4503, 2023. 3
- [EF23] Charilaos Efthymiou and Weiming Feng. On the mixing time of Glauber dynamics for the hard-core and related models on $G(n, d/n)$. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:17, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 3
- [Eft14] Charilaos Efthymiou. MCMC sampling colourings and independent sets of $G(n, d/n)$ near uniqueness threshold. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2014)*, SODA '14, pages 305–316, USA, 2014. Society for Industrial and Applied Mathematics. 2
- [Eft16] Charilaos Efthymiou. A simple algorithm for sampling colorings of $G(n, d/n)$ up to the Gibbs uniqueness threshold. *SIAM Journal on Computing*, 45(6):2087–2116, 2016. 2
- [Eft20] Charilaos Efthymiou. Deterministic counting of graph colourings using sequences of subgraphs. *Combinatorics, Probability and Computing*, 29(4):555–586, 2020. 2
- [EHŠV18] Charilaos Efthymiou, Thomas P. Hayes, Daniel Štefankovič, and Eric Vigoda. Sampling random colorings of sparse random graphs. In *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1759–1771, 2018. 2
- [FGYZ22] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. *ACM Trans. Algorithms*, 18(3), October 2022. 3
- [FK15] Alan Frieze and Michał Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015. 30
- [GŠV15] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6), December 2015. 2
- [Jon02] Johan Jonasson. Uniqueness of uniform random colorings of regular trees. *Statistics & Probability Letters*, 57(3):243–248, 2002. 2
- [JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. Spectral independence, coupling, and the spectral gap of the Glauber dynamics. *Information Processing Letters*, 177:106268, 2022. 3
- [LM11] Brendan Lucier and Michael Molloy. The Glauber dynamics for colorings of bounded degree trees. *SIAM Journal on Discrete Mathematics*, 25(2):827–853, 2011. 25
- [LP17] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2 edition, 2017. With contributions by Elizabeth L. Wilmer. 24, 26
- [LY26] Zhidan Li and Kuan Yang. Spatial mixing and deterministic approximate counting of multi-spin systems beyond bounded degree graphs. *arXiv: 2609.12352*, 2026. 5

- [Mar99] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In Pierre Bernard, editor, *Lectures on Probability Theory and Statistics: Ecole d'Été de Probabilités de Saint-Flour XXVII - 1997*, pages 93–191, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg. 25
- [MS10] Elchanan Mossel and Allan Sly. Gibbs rapidly samples colorings of $G(n, d/n)$. *Probability Theory and Related Fields*, 148(1):37–69, 2010. 2
- [Yin14] Yitong Yin. Spatial mixing of coloring random graphs. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *The 41st International Colloquium on Automata, Languages, and Programming (ICALP 2014)*, pages 1075–1086, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. 30
- [ZK07] Lenka Zdeborová and Florent Krzakala. Phase transitions in the coloring of random graphs. *Phys. Rev. E*, 76:031131, Sep 2007. 2, 5

APPENDIX A. SSM AND TID FOR LIST-COLORING INSTANCES

We record the list-coloring extension of Proposition 4.3 used in Section 4.

Lemma A.1. *Fix $\eta \in (0, 1)$, let $\Delta \geq 7/\eta^2$, and suppose that $q = (1 + \eta)\Delta$. Let T be a tree of maximum degree at most Δ , and let $\Phi = (T, L)$ be a feasible list-coloring instance satisfying*

$$(11) \quad |L(v)| - \deg_T(v) \geq q - \Delta \quad \text{for every } v \in V(T).$$

Then for any feasible pinning π on T , the Gibbs distribution μ_Φ^π exhibits strong spatial mixing and total influence decay with exponential decay rate $\exp(-\eta/4)$ and constants

$$C_{\text{SSM}} = \frac{2q}{(\eta\Delta)^{1/2}} \exp(1/\eta + \eta/4),$$

$$C_{\text{TID}} = \frac{q^3}{(\eta\Delta)^{3/2}(q-1)} \exp(3/\eta)$$

respectively.

Proof. Fix a feasible pinning π on $\Lambda \subseteq V(T)$. Let

$$T^\pi := T[V(T) \setminus \Lambda]$$

be the remaining free forest, and for every free vertex v let

$$L^\pi(v) := L(v) \setminus \{\pi(u) : u \in N_T(v) \cap \Lambda\}.$$

It clearly holds that

$$(12) \quad \begin{aligned} |L^\pi(v)| - \deg_{T^\pi}(v) &\geq |L(v)| - |N_T(v) \cap \Lambda| - (\deg_T(v) - |N_T(v) \cap \Lambda|) \\ &= |L(v)| - \deg_T(v) \\ &\geq q - \Delta. \end{aligned}$$

We now convert the list constraints as ordinary coloring constraints. For every free vertex v and every color $c \in [q] \setminus L^\pi(v)$, attach a new leaf $z_{v,c}$ to v and pin $z_{v,c}$ to color c . Denote the resulting augmented forest by $\hat{T}^\pi$. For every original free vertex v , by (12),

$$\deg_{\hat{T}^\pi}(v) = \deg_{T^\pi}(v) + q - |L^\pi(v)| \leq \Delta.$$

Meanwhile, all newly added vertices are leaves. Hence every component of $\hat{T}^\pi$ is a tree of maximum degree at most Δ .

Observe that conditional on the colors of the newly added leaves, the uniform distribution on proper q -colorings of $\hat{T}^\pi$, restricted to the original free vertices, is exactly μ_Φ^π . Apply [CLMM26, Proposition 8.8] to every component of $\hat{T}^\pi$. It gives strong spatial mixing and total influence

decay with rate $\exp(-\eta/4)$ and the displayed constants. Adding pinned leaves does not change distances between original vertices, and the added vertices never appear among the free vertices in the definitions of SSM or TID. Thus the same bounds hold for μ_{Φ}^{τ} on the original list-coloring instance, concluding this lemma. $\square$

APPENDIX B. POISSON ESTIMATIONS ON GALTON–WATSON TREES

This section provides a proof of Proposition 4.10. At beginning, we introduce the following properties of a Poisson random variable and its related variants.

Lemma B.1. *Let $K \sim \text{Poisson}(d+1)$ be a Poisson random variable of rate $d+1$. Recall that $D = \lfloor (1 + \eta/3)d \rfloor$. The following inequalities hold for sufficiently large $d > 1$.*

- **Poisson tail bound:** *there exists some constant $C_{\text{TAIL}} = C_{\text{TAIL}}(\eta) > 0$ such that*

$$p_{\mathcal{H}} := \mathbb{P}(K > D - 3) \leq e^{-C_{\text{TAIL}}d}.$$

- **Finite moments:** *for every integer $p \geq 2$, there exists some constant $C_{\text{MMT}} = C_{\text{MMT}}(\eta) > 0$ such that*

$$\|K\|_p, \|K+1\|_p, \|K \mid K > D-3\|_p \leq C_{\text{MMT}}(d+p).$$

Proof. The Poisson tail bound is a direct corollary of the Markov inequality: for every $x > d+1$,

$$\mathbb{P}(K \geq x) \leq \exp\left(-(d+1)\left(\frac{x}{d+1} \ln \frac{x}{d+1} - \frac{x}{d+1} + 1\right)\right).$$

For sufficiently large $d > 1$, $D-2 \geq (1 + \eta/5)(d+1)$ and thus

$$\mathbb{P}(K > D-3) \leq \mathbb{P}(K \geq D-2) \leq e^{-\eta^2 d/60}.$$

To show finite moments of random variables, we show the standard calculation for the moment generating function of K as

$$\mathbb{E}[e^K] = \exp((d+1)(e-1))$$

and therefore for every $x \geq 0$, by the Markov inequality,

$$\mathbb{P}(K > (d+1)(e-1) + t) \leq e^{-t}.$$

For the random variable $X = \max\{K - (d+1)(e-1), 0\}$, by the tail integral formula,

$$\mathbb{E}[X^p] = \int_0^\infty p x^{p-1} \mathbb{P}(X > x) dx \leq \int_0^\infty p x^{p-1} e^{-x} dx = p! \leq p^p.$$

Then by the Minkowski's inequality,

$$\|K\|_p \leq (e-1)(d+1) + p.$$

For the random variable $K+1$, the Minkowski's inequality gives that

$$\|K+1\|_p \leq 1 + \|K\|_p \leq 1 + (e-1)(d+1) + p.$$

To show the finite moments of $K \mid K > D-3$, we first pick $d > 1$ sufficiently large such that $K_0 := D-2 \geq (1 + \eta/5)(d+1)$. For integers $j, k > 0$, it holds that

$$\frac{\mathbb{P}(K \geq K_0 + j + k)}{\mathbb{P}(K \geq K_0 + j)} \leq \frac{(d+1)^k}{(K_0 + j + k) \cdots (K_0 + j + 1)} \leq \left(\frac{d+1}{K_0 + 1}\right)^k.$$

Hence we have

$$\mathbb{P}(K \geq K_0 + k \mid K \geq K_0) \leq \left(\frac{d+1}{K_0 + 1}\right)^k.$$

Consider the random variable $X = K - K_0$ conditional on $K \geq K_0$ and set $\theta = -\ln \frac{d+1}{K_0+1}$. For every $x \geq 0$, it holds that

$$\mathbb{P}(X > x) \leq \left(\frac{d+1}{K_0+1} \right)^x \leq e^{-\theta x}.$$

By the tail integral formula,

$$\mathbb{E}[X^p] = \int_0^\infty p x^{p-1} \mathbb{P}(X > x) dx \leq \int_0^\infty p x^{p-1} e^{-\theta x} dx = \frac{p!}{\theta^p} \leq \left(\frac{p}{\theta} \right)^p.$$

Then by the Minkowski's inequality,

$$\|K \mid K > D - 3\|_p = \|K \mid K \geq K_0\|_p \leq K_0 + \frac{p}{\theta} \leq \frac{6}{\eta}(d + p).$$

Pick $C_{\text{MMT}} = 6/\eta$ and we conclude the finite moments of these three random variables. $\square$

Based on the finite exponential moment of the offspring variable, we show a similar bound for the moment of the population through any fixed height.

Lemma B.2. *Let $\mathcal{T}$ be a Galton-Watson tree rooted at ϱ whose nodes have independent offspring counts distributed as a non-negative random variable U . Define N_j as the number nodes of depth j and $T_h := \sum_{j=0}^h N_j$. Then:*

- (1) *If $\mathbb{E}[e^{tU}] < \infty$ for some $t > 0$, then for every fixed integer $h \geq 0$, there exists $s_h > 0$ such that $\mathbb{E}[e^{s_h T_h}] < \infty$.*
- (2) *For every $p \geq 1$, if $\|U\|_p \leq B_p < \infty$, then $\|T_h\|_p \leq (1 + B_p)^{h+1}$.*

Proof. To show this lemma, observe that

$$\forall h \geq 1, \quad T_h \stackrel{d}{=} 1 + \sum_{i=1}^U T_{h-1}^{(i)}$$

where $T_{h-1}^{(i)}$ is an independent copy of T_{h-1} and is independent of U . To see this recursion, the root contributes itself. For each of its U children, it is an independent subtree of height $h-1$.

For simplicity, we denote by $M_h(s) := \mathbb{E}[e^{sT_h}]$ the moment generating function of T_h at s . By the recursion of T_h and the independence of random variables $T_{h-1}^{(i)}$,

$$\begin{aligned} M_h(s) &= e^s \mathbb{E} \left[\mathbb{E} \left[\prod_{i=1}^U e^{sT_{h-1}^{(i)}} \mid U \right] \right] \\ &= e^s \mathbb{E} [M_{h-1}(s)^U] \\ &= e^s \mathbb{E} [\exp(U \ln M_{h-1}(s))] . \end{aligned}$$

We prove the first argument by induction on h .

- **Boundary case:** $h = 0$. It holds that $T_0 = N_0 = 1$. Direct calculation on $M_0(s)$ tells that $M_0(s) = e^s < \infty$ for every $s > 0$.
- **Inductive case:** $h > 0$. Suppose that $M_{h-1}(\eta) < \infty$ for some $\eta > 0$. For every $s \in [0, \eta]$, $e^{sT_{h-1}} \leq e^{\eta T_{h-1}}$. Dominated convergence provides that as $s \downarrow 0$, $M_{h-1}(s) \rightarrow 1$ and consequently, $\ln M_{h-1}(s) \rightarrow 0$. Choose s_h sufficiently small to make

$$\ln M_{h-1}(s_h) \leq t_0.$$

Then by assumption $\mathbb{E}[e^{t_0 U}] < \infty$,

$$M_h(s_h) = e^{s_h} \mathbb{E} [\exp(U \ln M_{h-1}(s_h))] \leq e^{s_h} \mathbb{E} [e^{t_0 U}] < \infty.$$

We note here that the choice of s_h only depends on h .

For the bound of the ℓ_p -norm of T_h , by convexity of $x \mapsto x^p$,

$$\forall x_1, \dots, x_m \geq 0, \quad \left(\frac{1}{m} \sum_{i=1}^m x_i \right)^p \leq \frac{1}{m} \sum_{i=1}^m x_i^p.$$

Conditional on $U = m$,

$$\mathbb{E} \left[\left(\sum_{i=1}^U T_{h-1}^{(i)} \right)^p \middle| U = m \right] \leq m^{p-1} \sum_{i=1}^m \mathbb{E} \left[\left(T_{h-1}^{(i-1)} \right)^p \right] = m^p \mathbb{E} [T_{h-1}^p].$$

By law of total expectation,

$$\mathbb{E} \left[\left(\sum_{i=1}^U T_{h-1}^{(i)} \right)^p \right] \leq \mathbb{E} [U^p] \mathbb{E} [T_{h-1}^p]$$

which is equivalent to

$$\left\| \sum_{i=1}^U T_{h-1}^{(i)} \right\|_p \leq \|U\|_p \|T_{h-1}\|_p.$$

The Minkowski's inequality provides that

$$\forall h \geq 1, \quad \|T_h\|_p \leq 1 + B_p \|T_{h-1}\|_p$$

with the boundary condition $\|T_0\|_p = 1$. Solving this recursion, we conclude that $\|T_h\|_p \leq \sum_{i=0}^h B_p^i \leq (B_p + 1)^{h+1}$. $\square$

The next step is replacing high-degree marks by independent block marks to simplify the analysis. Set

$$L_{\text{BLK}} = L_{\text{BLK}}(\alpha) := \left\lceil \max \left\{ 3r, \frac{4 \ln 4}{\ln(1/\alpha)} \right\} \right\rceil.$$

It is clear to see $L_{\text{BLK}} = O_\eta((\ln d)^2)$. We partition the original tree into blocks containing levels $0, \dots, L_{\text{BLK}} - 1$ relative to each block root. The depth- L_{BLK} frontier of each block consists of the roots of its child blocks. Conditional on the existence of the root, a block reveals exactly the offspring of its internal nodes.

For a block, define $\mathcal{E}^{\text{mark}}$ as the event that it contains a node in $\mathcal{H}$. Set $\delta = \mathbb{P}(\mathcal{E}^{\text{mark}})$ for convenience. Independence of a node's offspring from its existence in its generation provides that

$$p_{\mathcal{H}} \leq \delta \leq \sum_{j=0}^{L_{\text{BLK}}-1} (d+1)^j p_{\mathcal{H}} \leq \exp(-C_{\text{TAIL}} d + O_\eta((\ln d)^3)).$$

Let S be the number of internal nodes plus the frontier size of a block. Independently sample S^+ with the conditional law $S \mid \mathcal{E}^{\text{mark}}$ and S^- as $S \mid \overline{\mathcal{E}^{\text{mark}}}$ and set $Z := S^+ + S^-$. Moreover, to choose the actual block data, independently pick a $\text{Bernoulli}(\delta)$ selector. Recursively place the selected child blocks in the first available slots of a node with Z children and fill extra slots independently by the same construction. This process embeds the actual block tree into a Galton-Watson tree with offspring Z and independent $\text{Bernoulli}(\delta)$ marks.

Proposition B.3. *For the random variable Z , there exists some $t > 0$ making $\mathbb{E}[e^{tZ}] < \infty$. Moreover, for every integer $p \geq 2$,*

$$\frac{\ln \mathbb{E}[(1+Z)^p]}{p} \leq C_\eta L_{\text{BLK}} \ln(d+p)$$

where C_η depends only on η .

Proof. Let S_h be the total population through depth h and set $M_h(t) := \mathbb{E}[e^{tS_h}]$. The branching recursion provides that

$$\forall h \geq 1, \quad M_h(t) = \exp(t + (d+1)(M_{h-1}(t) - 1))$$

with boundary condition $M_0(t) = e^t$. For every $t > 0$ and $h \geq 0$, this recursion implies that $M_h(t) < \infty$.

For $\mathcal{E}^{\text{mark}}$, it contains the event that the root of this block is in $\mathcal{H}$ while its complement contains the event that the root has no children. Therefore,

$$\mathbb{E}[e^{tS^+}] = \frac{\mathbb{E}[e^{tS} \mathbb{1}[\mathcal{E}^{\text{mark}}]]}{\delta} < \infty, \quad \mathbb{E}[e^{tS^-}] = \frac{\mathbb{E}[e^{tS} \mathbb{1}[\overline{\mathcal{E}^{\text{mark}}}}]]{1 - \delta} < \infty.$$

By independence of S^+ and S^- , $\mathbb{E}[e^{tZ}] = \mathbb{E}e^{tS^+} \mathbb{E}e^{tS^-} < \infty$.

Define a size-biased random variable K^* as

$$\mathbb{P}(K^* = k) = \frac{k\mathbb{P}(K = k)}{d+1}$$

and therefore it follows the law $1 + \text{Poisson}(d+1)$. Moreover, let K^+ be the conditional random variable with law $K \mid K > D-3$. By Lemma B.1,

$$\|K\|_p, \|K^*\|_p, \|K^+\|_p \leq C_{\text{MMT}}(d+p).$$

Choose $b_p = C(\eta)(d+p)$ such that

$$(13) \quad b_p \geq \max \left\{ 3, d+1, D+1, 1 + \max \left\{ \|K\|_p, \|K^*\|_p, \|K^+\|_p \right\} \right\}.$$

Observe that $\mathcal{E}^{\text{mark}}$ requires at least one high-degree node in the block:

$$\mathbb{1}[\mathcal{E}^{\text{mark}}] \leq \sum_{j=0}^{L_{\text{BLK}}-1} \sum_{\text{depth}(\omega)=j} \mathbb{1}[K_\omega > D-3].$$

Hence

$$(14) \quad \mathbb{E}[S^p \mathbb{1}[\mathcal{E}^{\text{mark}}]] \leq \sum_{j=0}^{L_{\text{BLK}}-1} \mathbb{E} \left[\sum_{\text{depth}(\omega)=j} S^p \mathbb{1}[K_\omega > D-3] \right].$$

Recall the law of the size-biased random variable K^* :

$$\mathbb{P}(K^* = k) = \frac{k\mathbb{P}(K = k)}{d+1}.$$

Hence we obtain that for every level j ,

$$\mathbb{E} \left[\sum_{\text{depth}(\omega)=j} S^p \mathbb{1}[K_\omega > D-3] \right] = (d+1)^j p_{\mathcal{H}} \mathbb{E}[\tilde{S}_j^p]$$

where $\tilde{S}_j$ is the total population through L_{BLK} in the following modified tree.

- The j ancestors on a distinguished path from root to a node of depth j have offspring law K^* .
- At each such ancestor, one uniform child continues the path.
- The endpoint has offspring law K^+ .
- Every other vertex has ordinary offspring law K .

Note that all offspring laws in the modified tree satisfy (13). By Lemma B.2,

$$\mathbb{E} [\tilde{S}_j^p] \leq b_p^{L_{\text{BLK}}}.$$

Plugging above arguments into (14), we have that

$$\mathbb{E} [S^p \mathbb{1} [\mathcal{E}^{\text{mark}}]] \leq p_{\mathcal{H}} \sum_{j=0}^{L_{\text{BLK}}-1} (d+1)^j b_p^{p L_{\text{BLK}}} \leq p_{\mathcal{H}} (b_p^p (d+1))^{L_{\text{BLK}}}$$

Since $\delta \geq p_{\mathcal{H}}$, for S^+ , it holds that

$$\mathbb{E} [(S^+)^p] = \frac{\mathbb{E} [S^p \mathbb{1} [\mathcal{E}^{\text{mark}}]]}{\delta} \leq L_{\text{BLK}} (d+1)^{L_{\text{BLK}}} b_p^{p L_{\text{BLK}}}.$$

Together with the choice of b_p , $\|S^+\|_p \leq b_p^{2L_{\text{BLK}}}$. For S^- , since every internal node has at most $D-3$ children,

$$S^- \leq \sum_{j=0}^{L_{\text{BLK}}} (D-3)^j \leq (1+D-3)^{L_{\text{BLK}}} \leq b_p^{L_{\text{BLK}}}$$

and thus $\|S^-\|_p \leq b_p^{L_{\text{BLK}}}$. By Minkowski's inequality,

$$\|1 + Z\|_p \leq 1 + \|S^+\|_p + \|S^-\|_p \leq b_p^{3L_{\text{BLK}}}.$$

Hence we conclude that $\ln \mathbb{E} [(1+Z)^p] \leq C_{\eta} p L_{\text{BLK}} \ln(d+p)$, which complete the proof. $\square$

A block is said to be bad if it is a marked block itself or there is a marked block within two block distance of it, and otherwise it is good. By $L_{\text{BLK}} \geq 3r$, every bad node in the original tree belongs to a bad block. Set

$$\beta := \alpha^{L_{\text{BLK}}}, \quad a := \frac{2(1+D^R)}{1-\alpha}.$$

Note that a good block has internal contribution at most a and transmits at most β times the largest frontier value, while the influence of a bad block is bounded by summing all internal costs and frontier influences. Therefore, the message $\mathcal{X}$ is dominated by a message on the block tree equipped with $\text{cost}(\omega) = a(1+Z_\omega)$ passed by the good rule $\beta \max$ and the bad rule sum. Let $\mathcal{X}^{\text{BLK}}$ be the message of the root on the block tree, and let $\mathcal{Y}^{\text{BLK}}$ be the message with an additional marked parent outside the tree. Therefore

$$\mathcal{X} \leq \mathcal{X}^{\text{BLK}}, \quad \mathcal{Y} \leq \mathcal{Y}^{\text{BLK}}.$$

We provide the following proposition to $\mathcal{X}^{\text{BLK}}$ and $\mathcal{Y}^{\text{BLK}}$.

Proposition B.4. *Assume all definitions and notations above. Let $m = \mathbb{E} [Z]$, N_h be the generation- h population, $C_h := \sum_{\text{depth}(\omega) < h} a(1+Z_\omega)$ and $b_h := \mathbb{E} [N_h \cdot \mathbb{1} [\mathcal{E}]]$. Let $\mathcal{E}$ be the event that there is a mark at depth zero, one or two. Then*

$$p_{\mathcal{E}} := \mathbb{P} (\mathcal{E}) \leq \delta(1+m+m^2).$$

Suppose that there is an integer $P \geq 2$ satisfying that

$$(15) \quad \beta m^{1/P} \leq \frac{1}{4},$$

$$(16) \quad p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P} \leq \frac{1}{4},$$

$$(17) \quad p_{\mathcal{E}}^{1/(2P)} \|N_2\|_P \|N_3\|_{2P} \leq \frac{1}{16}.$$

Then the matrix

$$B_P := \begin{pmatrix} \beta^{P+1}m & b_3 \\ m^2 & b_3 \end{pmatrix}$$

has spectral radius less than 1 and for some $s > 0$, $\mathbb{E}[\exp(s\mathcal{X}^{\text{BLK}})] < \infty$ and $\mathbb{E}[\exp(s\mathcal{Y}^{\text{BLK}})] < \infty$.

Proof. To show $p_{\mathcal{E}} \leq \delta(1 + m + m^2)$, note that conditional on the tree, the probability of $\mathcal{E}$ is at most $\delta(N_0 + N_1 + N_2)$. Taking expectations provides the desired upper bound.

For $h \geq 1$, let X_h and Y_h denote the messages restricted to depths $0, \dots, h-1$ with zero frontier values. For notational convenience, set

$$X_h = Y_h = 0 \quad \text{for every integer } h \leq 0.$$

Write

$$M_h^X(s) := \mathbb{E}[e^{sX_h}], \quad M_h^Y(s) := \mathbb{E}[e^{sY_h}].$$

Consequently, $M_h^X(s) = M_h^Y(s) = 1$ for $h \leq 0$, and the non-negativity of all costs and update rules gives

$$X_h \uparrow \mathcal{X}^{\text{BLK}}, \quad Y_h \uparrow \mathcal{Y}^{\text{BLK}} \quad \text{as } h \rightarrow \infty.$$

We claim that X_h and Y_h have positive exponential moments. In fact, by induction, it is direct to see $X_h \leq C_h$ and $Y_h \leq C_h$. An argument similar to the proof of Lemma B.2 shows that, for every finite $h \geq 0$, there exists some $s_h > 0$ such that $\mathbb{E}[\exp(s_h X_h)]$ and $\mathbb{E}[\exp(s_h Y_h)]$ are finite.

Let $\mathcal{F}_3$ be the sigma-field generated by the block tree through generation three, together with the marks in generations zero, one, and two. Then $\mathcal{E} \in \mathcal{F}_3$, and, conditional on $\mathcal{F}_3$, the marked descendant trees rooted at generation three are independent unconditional copies of the original marked Galton–Watson tree.

On $\bar{\mathcal{E}}$, couple the conditioned marked tree with an unconditional marked tree by keeping all offspring variables and independently resampling the marks that were forced to be zero. The message is monotone in the collection of marks: changing a good node into a bad node replaces a $\beta \max$ update by a sum update, and $\beta \max_i x_i \leq \sum_i x_i$ for non-negative inputs. Therefore, using fresh unconditional copies on the right-hand side,

$$\begin{aligned} X_h &\preceq a(1 + Z) + \beta \max_{1 \leq i \leq Z} X_{h-1}^{(i)}, \\ Y_h &\preceq C_2 + \sum_{i=1}^{N_2} X_{h-2}^{(i)} \quad \text{on } \bar{\mathcal{E}}. \end{aligned}$$

The copies belonging to distinct descendant branches are independent conditional on the revealed prefix.

On $\mathcal{E}$, expand the bad-rule recursion through the first three generations. Conditional on $\mathcal{F}_3$, the subtrees below the generation-three vertices are independent and unconditional. Adding a marked parent above each such subtree can only increase its message. It follows that

$$X_h, Y_h \preceq C_3 + \sum_{i=1}^{N_3} Y_{h-3}^{(i)}.$$

Taking exponential moments conditional on the revealed prefix and then averaging gives the following bounds. In the first term of each bound, dropping the indicator of $\bar{\mathcal{E}}$ can only increase every Taylor

coefficient. Therefore, coefficient-wise,

$$\begin{aligned} M_h^X(s) &\preceq \mathbb{E} \left[e^{sa(1+Z)} (1 + Z (M_{h-1}^X(\beta s) - 1)) \right] + \mathbb{E} \left[\mathbb{1}[\mathcal{E}] e^{sC_3} (M_{h-3}^Y(s))^{N_3} \right], \\ M_h^Y(s) &\preceq \mathbb{E} \left[e^{sC_2} (M_{h-2}^X(s))^{N_2} \right] + \mathbb{E} \left[\mathbb{1}[\mathcal{E}] e^{sC_3} (M_{h-3}^Y(s))^{N_3} \right]. \end{aligned}$$

For every finite integer $M \geq 0$, define

$$H_M := \max_{h=0,\dots,M} \|X_h\|_P, \quad K_M := \max_{h=0,\dots,M} \|Y_h\|_P.$$

We upper bound H_M and K_M by the above decomposition. It is clear to see

$$\mathbb{E} \left[\left(\max_{i=1,\dots,Z} X_{h-1}^{(i)} \right)^P \right] \leq \mathbb{E} \left[\sum_{i=1}^Z \left(X_{h-1}^{(i)} \right)^P \right] \leq m H_M^P.$$

For the sum rule, convexity provides that

$$\mathbb{E} \left[\left(\sum_{i=1}^N W_i \right)^P \right] \leq N^P \max_{i=1,\dots,N} \mathbb{E} [W_i^P].$$

Together with the decomposition, it tells that

$$\begin{aligned} \left\| \mathbb{1}[\mathcal{E}] \sum_{i=1}^{N_3} \mathcal{Y}_{h-3}^{(i)} \right\|_P^P &= \mathbb{E} \left[\mathbb{1}[\mathcal{E}] \mathbb{E} \left[\left(\sum_{i=1}^{N_3} \mathcal{Y}_{h-3}^{(i)} \right)^P \middle| \text{revealed depth-three nodes} \right] \right] \\ &\leq K_M^P \mathbb{E} [\mathbb{1}[\mathcal{E}] N_3^P] \\ &= K_M^P \|\mathbb{1}[\mathcal{E}] N_3\|_P^P. \end{aligned}$$

By Cauchy-Schwarz inequality,

$$\|\mathbb{1}[\mathcal{E}] N_3\|_P \leq p_{\mathcal{E}}^{1/(2P)} \mathbb{E} [N_3^{2P}]^{1/(2P)}.$$

Similarly,

$$\left\| \sum_{i=1}^{N_2} X_{h-2}^{(i)} \right\|_P \leq H_M \|N_2\|_P.$$

Minkowski's inequality gives that

$$\begin{aligned} H_M &\leq c_0 + \beta m^{1/P} H_M + K_M \cdot p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P}, \\ K_M &\leq c_1 + \|N_2\|_P H_M + K_M \cdot p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P} \end{aligned}$$

where $c_0 = \|a(1+Z)\|_P + \|C_3\|_P$ and $c_1 = \|C_2\|_P + \|C_3\|_P$ are both finite positive terms only depending on fixed-height weights. By assumption,

$$K_M \leq \frac{4}{3} (c_1 + \|N_2\|_P H_M) \quad \text{and thus} \quad H_M \leq c_0 + \frac{4}{3} c_1 \cdot p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P} + \frac{1}{3} H_M$$

meaning

$$\sup_{M \geq 0} H_M < \infty, \quad \sup_{M \geq 0} K_M < \infty.$$

By monotonicity of ℓ_p -norm, all moments mentioned in this proof of orders $0, \dots, P$ are also uniformly bounded.

For higher moments, choose polynomials f_0, g_0 of degree at most P with constant coefficient 1 and non-negative coefficients satisfying

$$[s^i]f_0(s) \geq \frac{\mathbb{E}[X_h^i]}{i!}, \quad [s^i]g_0(s) \geq \frac{\mathbb{E}[Y_h^i]}{i!}$$

for every $h \geq 0$. The existence of f_0 and g_0 comes from the uniform upper bound for H_M and K_M . Our aim is to find two convergent series u, v with non-negative coefficients and no terms of degree at most P such that

$$\mathbb{E}[e^{sX_h}] \preceq f_0(s) + u(s), \quad \mathbb{E}[e^{sY_h}] \preceq g_0(s) + v(s)$$

where $\preceq$ means coefficient-wise comparison.

We first state two useful observations.

- For every $x_1, \dots, x_N \geq 0$,

$$e^{s\beta \max_{i=1}^N x_i} \preceq 1 + \sum_{i=1}^N (e^{s\beta x_i} - 1).$$

- Since u and v start at degree $P+1$,

$$u(\beta s) \preceq \beta^{P+1}u(s), \quad v(\beta s) \preceq \beta^{P+1}v(s).$$

For scalar variables s, u and v , define

$$\begin{aligned} \Psi_1(s, u, v) &:= \mathbb{E} \left[e^{sa(1+Z)} (1 + Z(f_0(\beta s) - 1 + \beta^{P+1}u)) \right] + \mathbb{E} [\mathbb{1}[\mathcal{E}] e^{sC_3} (g_0(s) + v)^{N_3}], \\ \Psi_2(s, u, v) &:= \mathbb{E} \left[e^{sC_2} (f_0(s) + u)^{N_2} \right] + \mathbb{E} [\mathbb{1}[\mathcal{E}] e^{sC_3} (g_0(s) + v)^{N_3}]. \end{aligned}$$

We first verify that Ψ_1, Ψ_2 are analytic around the origin $(0, 0, 0)$. By Proposition B.3, there exists some $t_0 > 0$ such that $\mathbb{E}[e^{t_0 Z}] < \infty$. Consequently, the population through any fixed depth has a positive exponential moment. Indeed, for every $h \geq 0$, define

$$M_h(t) := \mathbb{E}[\exp(t(N_0 + \dots + N_h))].$$

The branching recursion for the Galton–Watson tree provides that

$$M_0(t) = e^t, \quad M_h(t) = e^t \mathbb{E}[M_{h-1}(t)^Z].$$

If M_{h-1} is finite near 0, then $M_{h-1}(t) \rightarrow 1$ as $t \downarrow 0$. We pick t sufficiently small such that $\ln M_{h-1}(t) \leq t_0$ which makes $M_h(t)$ finite. It is enough to pick t depending only on h and the offspring distribution Z . Specifically, let $T := N_0 + N_1 + N_2 + N_3$ be the population from depth zero to three. There exists a constant $\theta > 0$ such that

$$M_3(\theta) = \mathbb{E}[e^{\theta T}] < \infty.$$

Observe that $C_2 = a(N_0 + 2N_1 + N_2)$ and $C_3 = a(N_0 + 2N_1 + 2N_2 + N_3)$. Then it is clear that $a(1+Z) \leq aT$ and $C_2, C_3 \leq 2aT$. Choose $L \geq 1$ such that for complex $|s| \leq 1$,

$$|f_0(s) - 1| \leq L|s|, \quad |g_0(s) - 1| \leq L|s|.$$

Pick the radius R as

$$R := \min \left\{ \frac{1}{2}, \frac{\theta}{2(2a + L + 1)} \right\}.$$

Write

$$\begin{aligned} G_1(s, u, v) &:= e^{sa(1+Z)} (1 + Z(f_0(\beta s) - 1 + \beta^{P+1}u)), \\ G_2(s, u, v) &:= e^{sC_2} (f_0(s) + u)^{N_2}, \\ G_3(s, u, v) &:= 1[\mathcal{E}] e^{sC_3} (g_0(s) + v)^{N_3} \end{aligned}$$

and thus $\Psi_1 = \mathbb{E}[G_1 + G_3]$ and $\Psi_2 = \mathbb{E}[G_2 + G_3]$. We show that for every $s, u, v \in \mathbb{C}$ such that $|s|, |u|, |v| \leq R < 1$, G_1, G_2 and G_3 are upper bounded by $e^{(2a+L+1)RT}$.

For $G_1(s, u, v)$, since $\beta \in (0, 1)$ and $|s|, |u| \leq R \leq 1$,

$$|f_0(\beta s) - 1 + \beta^{P+1}u| \leq |f_0(\beta s) - 1| + \beta^{P+1}|u| \leq L|\beta s| + |u| \leq (L+1)R$$

By the inequality $1 + x \leq e^x$ on $[0, \infty)$,

$$|G_1(s, u, v)| \leq e^{aR(1+Z)} (1 + Z(L+1)R) \leq e^{(a+L+1)R(1+Z)} \leq e^{(2a+L+1)RT}.$$

For $G_2(s, u, v)$, by triangle inequality,

$$|f_0(s) + u| \leq 1 + |f_0(s) - 1| + |u| \leq 1 + (L+1)R \leq e^{(L+1)R}.$$

Therefore,

$$|G_2(s, u, v)| \leq e^{RC_2} |f_0(s) + u|^{N_2} \leq e^{2aRT} e^{(L+1)RN_2} \leq e^{(2a+L+1)RT}.$$

A similar argument also holds for $G_3(s, u, v)$ and therefore $|G_3(s, u, v)| \leq e^{(2a+L+1)RT}$.

Now we are ready to show analyticity of Ψ_1 and Ψ_2 . Fix any one of G_1, G_2 and G_3 and denote it by G . Observe that for every realization of the tree ω , G is an entire function consisting of the product of the exponential and polynomials. Expand its Taylor series:

$$G(\omega; s, u, v) = \sum_{i,j,k \geq 0} c_{i,j,k}(\omega) s^i u^j v^k.$$

Cauchy's coefficient formula gives

$$c_{i,j,k}(\omega) = \frac{1}{(2\pi i)^3} \int_{|\zeta_1|=R} \int_{|\zeta_2|=R} \int_{|\zeta_3|=R} \frac{G(\omega; \zeta_1, \zeta_2, \zeta_3)}{\zeta_1^{i+1} \zeta_2^{j+1} \zeta_3^{k+1}} d\zeta_3 d\zeta_2 d\zeta_1.$$

The above modular bound for G shows that

$$|c_{i,j,k}(\omega)| \leq e^{\theta T(\omega)/2} R^{-(i+j+k)}$$

and each coefficient is integrable. Therefore,

$$\mathbb{E}[G(s, u, v)] = \sum_{i,j,k \geq 0} \mathbb{E}[c_{i,j,k}] s^i u^j v^k.$$

This power series converges absolutely and uniformly over every smaller polydisc

$$\{(s, u, v) : |s| \leq r, |u| \leq r, |v| \leq r\} \quad 0 \leq r < R.$$

Therefore $\mathbb{E}[G]$ is analytic around the origin. This leads to the claim that Ψ_1 and Ψ_2 are analytic around $(0, 0, 0)$. Moreover, since $f_0(\beta s) - 1$ has non-negative coefficients, every multivariate Taylor coefficient of Ψ_1 and Ψ_2 is non-negative.

For a power series H , let T_P be the Taylor contraction of H through the degree P :

$$\mathsf{T}_P H(s) = \sum_{j=0}^P [s^j] H(s) s^j.$$

Define

$$F(s, u, v) := \begin{pmatrix} \Psi_1(s, u, v) \\ \Psi_2(s, u, v) \end{pmatrix} - \mathsf{T}_P \begin{pmatrix} \Psi_1(s, 0, 0) \\ \Psi_2(s, 0, 0) \end{pmatrix}.$$

Differentiate F on (u, v) at the origin, and we obtain that

$$D_{(u,v)}F|_{(0,0,0)} = B_P = \begin{pmatrix} \beta^{P+1}m & b_3 \\ m^2 & b_3 \end{pmatrix}$$

By monotonicity of ℓ_q -norms and Cauchy-Schwarz inequality,

$$b_3 = \|N_3 \cdot \mathbb{1}[\mathcal{E}]\|_1 \leq \|N_3 \cdot \mathbb{1}[\mathcal{E}]\|_P \leq p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P} \leq \frac{1}{4}.$$

For $z = (x, y)^{\mathsf{T}} \in \mathbb{C}^2$, define the weighted norm

$$\|z\|_w = \max \left\{ |x|, \frac{|y|}{1 + 4m^2} \right\}.$$

The operator norm of B_P is

$$\|B_P\|_w = \max \left\{ \beta^{P+1}m + b_3(1 + 4m^2), \frac{m^2}{1 + 4m^2} + b_3 \right\}.$$

By assumption, $\|B_P\|_w \leq \frac{9}{16} < 1$. Since F is analytic and its derivative is continuous, we can choose $r_0, \varepsilon > 0$ such that F is analytic on a neighborhood of

$$\{(s, z) : |s| \leq r_0, \|z\|_w \leq \varepsilon\}$$

and

$$\|D_z F(s, z)\|_w \leq \frac{3}{4}$$

through this set. Integrating the derivative along the segment between z and z' gives that

$$\|F(s, z) - F(s, z')\|_w \leq \frac{3}{4} \|z - z'\|_w.$$

On the other hand, $F(s, 0) = O(s^{P+1})$. Thus after reducing r_0 if necessary, there is a constant $K \geq 1$ such that

$$|F(s, 0)|_w \leq K|s|^{P+1}, \quad \forall |s| \leq r_0.$$

Choose $0 < r \leq r_0$ small such that $Kr^{P+1} \leq \varepsilon/4$. Therefore, for every $|s| \leq r$ and $\|z\|_w \leq \varepsilon$,

$$\begin{aligned} \|F(s, z)\|_w &\leq \|F(s, z) - F(s, 0)\|_w + \|F(s, 0)\|_w \\ &\leq \frac{3}{4} \|z\|_w + K|s|^{P+1} \leq \frac{3}{4} \varepsilon + \frac{1}{4} \varepsilon = \varepsilon. \end{aligned}$$

Hence, for every $|s| \leq r$, $F(s, \cdot)$ maps a closed ball into itself and is a contraction. Define

$$U_0(s) = (0, 0), \quad U_{k+1}(s) = F(s, U_k(s)).$$

By Banach's fixed-point theorem, the iteration goes to a limit (u, v) . Uniform convergence on compact subsets preserves analyticity. Therefore, u, v are analytic on $|s| < r$.

Now we verify the required properties of their Taylor coefficients. Write

$$\Psi_i(s, u, v) = \sum_{j,k,\ell \geq 0} c_{j,k,\ell}^{(i)} s^j u^k v^\ell, \quad c_{j,k,\ell}^{(i)} \geq 0.$$

Therefore

$$F_i(s, u, v) = \sum_{j \geq P+1} c_{j,0,0}^{(i)} s^j + \sum_{j,k,\ell \geq 0, k+\ell \geq 1} c_{j,k,\ell}^{(i)} s^j u^k v^\ell.$$

Starting from $U_0 = (0, 0)$, induction shows that every component of U_k has non-negative coefficients and vanishes through degree P . These properties pass to the limit coefficient by coefficient. In fact, for any $0 < t < r$, Cauchy's coefficient formula and uniform convergence on $|s| = t$ provides that

$$[s^n]U(s) = \frac{1}{2\pi i} \int_{|s|=t} \frac{U(s)}{s^{n+1}} ds = \lim_{k \rightarrow \infty} [s^n]U_k(s).$$

Thus

$$u(s) = \sum_{n=P+1}^{\infty} u_n s^n, \quad v(s) = \sum_{n=P+1}^{\infty} v_n s^n$$

and both series converge on $|s| < r$.

Since $u(s)$ and $v(s)$ vanish through degree P , every monomial in $\Psi_i(s, u(s), v(s))$ containing at least one factor of $u(s)$ or $v(s)$ has degree at least $P+1$ in s . Therefore,

$$(18) \quad \mathsf{T}_P \Psi_i(s, u(s), v(s)) = \mathsf{T}_P \Psi_i(s, 0, 0), \quad i = 1, 2.$$

Consequently, the fixed-point equation and (18) imply

$$(19) \quad u(s) = (I - \mathsf{T}_P) \Psi_1(s, u(s), v(s)),$$

$$(20) \quad v(s) = (I - \mathsf{T}_P) \Psi_2(s, u(s), v(s)).$$

We now prove by strong induction on $h \in \mathbb{Z}$ that

$$(21) \quad M_h^X(s) \preceq f_0(s) + u(s),$$

$$(22) \quad M_h^Y(s) \preceq g_0(s) + v(s).$$

For $h \leq 0$, both moment generating functions equal 1, so the claim follows from the non-negativity of the coefficients of f_0, g_0, u , and v .

Fix $h \geq 1$ and suppose that the claim holds for every integer $k < h$. Using the coefficient-wise recursions above, the induction hypothesis, and $u(\beta s) \preceq \beta^{P+1}u(s)$, we obtain

$$M_h^X(s) \preceq \Psi_1(s, u(s), v(s)),$$

$$M_h^Y(s) \preceq \Psi_2(s, u(s), v(s)).$$

By the definition of f_0 and g_0 ,

$$\mathsf{T}_P M_h^X(s) \preceq f_0(s),$$

$$\mathsf{T}_P M_h^Y(s) \preceq g_0(s).$$

For the higher-order coefficients, (19) and (20) give

$$(I - \mathsf{T}_P) M_h^X(s) \preceq (I - \mathsf{T}_P) \Psi_1(s, u(s), v(s)) = u(s),$$

$$(I - \mathsf{T}_P) M_h^Y(s) \preceq (I - \mathsf{T}_P) \Psi_2(s, u(s), v(s)) = v(s).$$

Adding the low- and high-degree parts proves (21) and (22).

For every fixed $s \in (0, r)$, the two majorants are finite. Since $X_h \uparrow \mathcal{X}^{\text{BLK}}$ and $Y_h \uparrow \mathcal{Y}^{\text{BLK}}$, the monotone convergence theorem gives

$$\mathbb{E} \left[e^{s \mathcal{X}^{\text{BLK}}} \right] = \lim_{h \rightarrow \infty} M_h^X(s) \leq f_0(s) + u(s) < \infty,$$

$$\mathbb{E} \left[e^{s \mathcal{Y}^{\text{BLK}}} \right] = \lim_{h \rightarrow \infty} M_h^Y(s) \leq g_0(s) + v(s) < \infty. \quad \square$$

Proof of Proposition 4.10. We verify (15), (16) and (17) to show that $\mathcal{X}^{\text{BLK}}$ and $\mathcal{Y}^{\text{BLK}}$ have finite positive exponential moments and therefore

$$\mathbb{E} [e^{t_0 \mathcal{X}}] < \infty, \quad \mathbb{E} [e^{t_0 \mathcal{Y}}] < \infty.$$

for some constant $t_0 > 0$. Note that $\ln m = O_\eta(L_{\text{BLK}} \ln d)$ and by convexity,

$$\|N_j\|_p \leq \|Z\|_p^j.$$

Therefore, for fixed $j \geq 0$, $\ln \|N_j\|_p = O_\eta(L_{\text{BLK}} \ln(d + p))$. Choose $P = \lceil C_{\eta, \alpha} \ln(d + 2) \rceil$ for a sufficiently large constant. It is easy to verify

$$\beta m^{1/P} \leq \exp\left(-\frac{1}{2} \ln(1/\alpha) L_{\text{BLK}}\right) \leq \frac{1}{4}.$$

Meanwhile,

$$\ln p_{\mathcal{E}} \leq -c_\eta d + O_\eta(L_{\text{BLK}} \ln d).$$

This gives that

$$\max \{p_{\mathcal{E}}^{1/(2P)} \|N_3\|_{2P}, p_{\mathcal{E}}^{1/(2P)} \|N_2\|_P \|N_3\|_{2P}\} \leq \exp\left(-\Omega_{\eta, \alpha}\left(\frac{d}{\ln d}\right) + O_{\eta, \alpha}((\ln d)^3)\right)$$

which is bounded by $1/16$ for sufficiently large d . Then by Proposition B.4, the finite positive exponential moment bound is concluded for some constant $t_0 > 0$. $\square$